

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 15:57:50 ; Search time 71.3326 Seconds
(without alignments)
4515.528 Million cell updates/sec

Title: US-10-092-390-2
Perfect score: 6744
Sequence: 1 MVISLNSCLSFICLLCHWI.....SSPKQEDSGGSSSNSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		%					Description
Result	Query						
No.	Score	Match	Length	DB	ID		
1	6744	100.0	1140	5	AAE27985	Aae27985	Human EGF
2	6744	100.0	1140	7	ADD18688	Add18688	Human dis
3	6744	100.0	1192	7	ADE71305	Ade71305	Novel hum
4	3601	53.4	586	5	AAE27986	Aae27986	Human EGF
5	3107.5	46.1	878	4	ABG08033	Abg08033	Novel hum
6	2506.5	37.2	1050	4	AAB66267	Aab66267	Human TAN
7	2482.5	36.8	994	5	AAG79417	Aag79417	CADHP-6,
8	1909	28.3	636	4	AAB66269	Aab66269	Rat TANGO
9	1897	28.1	1350	6	ADA21141	Ada21141	Human sec

10	1879	27.9	1577	6	ABJ37904	Abj37904	NOVX	prot
11	1874.5	27.8	1261	7	ADD78227	Add78227	Human	CGD
12	1860	27.6	349	6	ABP75770	Abp75770	Human	sec
13	1858.5	27.6	1450	6	ABJ37901	Abj37901	NOVX	prot
14	1847.5	27.4	739	6	ABU03489	Abu03489	Angiogene	
15	1770	26.2	1403	6	ABJ37903	Abj37903	NOVX	prot
16	1761.5	26.1	1398	6	ABJ37900	Abj37900	NOVX	prot
17	1761.5	26.1	1404	6	ABJ37899	Abj37899	NOVX	prot
18	1708	25.3	1097	6	ADA21140	Ada21140	Human	sec
19	1522	22.6	384	4	AAG75479	Aag75479	Human	col
20	1466	21.7	321	4	ABG27639	Abg27639	Novel	hum
21	1272	18.9	466	4	ABG22559	Abg22559	Novel	hum
22	1252	18.6	434	4	ABB66756	Abb66756	Drosophil	
23	1241.5	18.4	762	4	ABG08032	Abg08032	Novel	hum
24	1192	17.7	474	4	AAy72715	Aay72715	HFICU08	c
25	1169	17.3	269	4	ABG08031	Abg08031	Novel	hum
26	1169	17.3	269	6	ABO00812	Abo00812	Polypepti	
27	1169	17.3	311	6	ABO00512	Abo00512	Novel	hum
28	1034.5	15.3	2444	5	ABB07821	Abb07821	Constitut	
29	1034.5	15.3	2556	2	AAO27066	Aao27066	Human	Not
30	1034.5	15.3	2556	6	ABG70518	Abg70518	Human	pol
31	1034.5	15.3	2556	6	AAG79773	Aag79773	Human	Not
32	1034.5	15.3	2556	6	ABP72571	Abp72571	Human	Not
33	1034.5	15.3	2556	6	ABR61830	Abr61830	Human	Not
34	1034.5	15.3	2556	7	ABR61759	Abr61759	Human	Not
35	1024	15.2	2531	7	ADE63713	Ade63713	Rat	Prote
36	1024	15.2	2531	7	ADE63705	Ade63705	Rat	Prote
37	1024	15.2	2531	7	ADE63709	Ade63709	Rat	Prote
38	1024	15.2	2531	7	ADE63701	Ade63701	Rat	Prote
39	1018.5	15.1	661	6	ABU11760	Abu11760	Human	MDD
40	1014.5	15.0	1473	5	AAE18208	Aae18208	Human	MOL
41	1014.5	15.0	1473	7	ADD18194	Add18194	Human	mol
42	1014.5	15.0	2471	2	AAO27065	Aao27065	Human	Not
43	1014.5	15.0	2471	6	AAG79774	Aag79774	Human	Not
44	1014.5	15.0	2471	6	ABP72572	Abp72572	Human	Not
45	1014.5	15.0	2471	6	ABR61831	Abr61831	Human	Not

ALIGNMENTS

RESULT 1
AAE27985
ID AAE27985 standard; protein; 1140 AA.
XX
AC AAE27985;
XX
DT 27-JAN-2003 (first entry)
XX
DE Human EGF-family protein #1.
XX
KW Human; EGF-family protein; novel human protein; NHP; drug discovery;
KW restriction fragment length polymorphism analysis; forensic biology;
KW toxicity; infectious disease; biological disorder; medical disorder;
KW mental disorder; gene therapy.
XX
OS Homo sapiens.

XX
 PN WO200272611-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 06-MAR-2002; 2002WO-US007477.
 XX
 PR 12-MAR-2001; 2001US-0275013P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Yu X, Miranda M;
 XX
 DR WPI; 2002-723315/78.
 DR N-PSDB; AAD46318.
 XX
 PT New novel human nucleic acids useful for e.g. identifying protein coding
 PT sequences and mapping unique genes to a particular chromosome, as DNA
 PT markers for restriction fragment length polymorphism analysis, or in
 PT forensic biology.
 XX
 PS Claim 2; Page 37-40; 42pp; English.
 XX
 CC The present sequence is EGF-family protein, a novel human protein (NHP).
 CC The NHP sequences are useful for mapping unique genes to a particular
 CC chromosome; as DNA markers for restriction fragment length polymorphism
 CC analysis; in forensic biology; in defining and monitoring both drug
 CC action and toxicity; in identifying, selecting and validating novel
 CC molecular targets for drug discovery; in microarrays or other assay
 CC formats to screen collections of genetic material from patients who have
 CC a particular medical condition. The NHP peptides, fusion proteins,
 CC antibodies, antagonists and agonists can be used for detecting mutant
 CC NHPs or inappropriately expressed NHPs for the diagnosis of disease; for
 CC screening drugs for treatment of symptomatic or phenotypic manifestations
 CC of perturbing the normal function of NHP in the body and to treat
 CC diseases including infectious, mental, biological, or medical diseases or
 CC disorders. They are also used in gene therapy
 XX
 SQ Sequence 1140 AA;

Query Match 100.0%; Score 6744; DB 5; Length 1140;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

 Qy 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

 Qy 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180

 Qy 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC 240

Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSGLCEAGFAGERCEARLCPEGLY	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSGLCEAGFAGERCEARLCPEGLY	360
Qy	361	GIKCDKRCPCHLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Db	361	GIKCDKRCPCHLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Qy	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYLNCAE	540
Db	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYLNCAE	540
Qy	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGIC	600
Db	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGIC	600
Qy	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVC	660
Db	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVC	660
Qy	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Db	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Qy	721	GAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKGDCALICQCQNGADCDHISGQCTCRTGF	780
Db	721	GAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKGDCALICQCQNGADCDHISGQCTCRTGF	780
Qy	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGVIIIVGNLN	840
Db	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGVIIIVGNLN	840
Qy	841	SLSRTSTALPADSYQIGAIAGIIILVLVVLFLALFIIYRHKQKGKESSMPAVTYTPAMR	900
Db	841	SLSRTSTALPADSYQIGAIAGIIILVLVVLFLALFIIYRHKQKGKESSMPAVTYTPAMR	900
Qy	901	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFV	960
Db	901	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFV	960
Qy	961	NLKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN	1020
Db	961	NLKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN	1020
Qy	1021	CSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT	1080

Db 1021 CSLSSSENFPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT 1080

Qy 1081 VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE 1140
 |||||

Db 1081 VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE 1140

RESULT 2

ADD18688

ID ADD18688 standard; protein; 1140 AA.

XX

AC ADD18688;

XX

DT 15-JAN-2004 (first entry)

XX

DE Human disease related protein SeqID119.

XX

KW human; disease state; cytostatic; antiinflammatory; ophthalmological;

KW antiarteriosclerotic; vulnerary; gene therapy;

KW hypoxia-regulated condition; tumourigenesis; angiogenesis; apoptosis;

KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;

KW glucose transportation; catecholamine synthesis; iron transport;

KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;

KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;

KW inflammatory condition; wound healing.

XX

OS Homo sapiens.

XX

PN WO2003018621-A2.

XX

PD 06-MAR-2003.

XX

PF 23-AUG-2002; 2002WO-GB003892.

XX

PR 23-AUG-2001; 2001GB-00020558.

PR 05-OCT-2001; 2001GB-00024037.

XX

PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX

PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX

DR WPI; 2003-290046/28.

DR N-PSDB; ADD18689.

XX

PT New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.

XX

PS Claim 25; SEQ ID NO 119; 424pp; English.

XX

CC This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,

CC such as tumourigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of
 CC a disease related protein of the invention.

XX

SQ Sequence 1140 AA;

Query Match 100.0%; Score 6744; DB 7; Length 1140;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	1	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Qy	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Db	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEGLY	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEGLY	360
Qy	361	GIKCDKRCPCHELENTSHCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Db	361	GIKCDKRCPCHELENTSHCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGV	480
Qy	481	DCSIRCPSTGWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAE	540
Db	481	DCSIRCPSTGWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAE	540
Qy	541	RCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGIC	600
Db	541	RCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGIC	600
Qy	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVC	660

Db	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQC	VHSSGPCHHITGLCDCLPGFTGALCNEVC	660
Qy	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN		720
Db	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN		720
Qy	721	GAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF		780
Db	721	GAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF		780
Qy	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTTCYCSPGWKGARCDQAGVIIVGNLN		840
Db	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTTCYCSPGWKGARCDQAGVIIVGNLN		840
Qy	841	SLSRTSTALPADSYQIGAIAGIIILVLVVLFLALFIIYRHKQKGKESSMPAVTYTPAMR		900
Db	841	SLSRTSTALPADSYQIGAIAGIIILVLVVLFLALFIIYRHKQKGKESSMPAVTYTPAMR		900
Qy	901	VVNADYTISGTLPHSNGGNANSHYFTNP	SYHTLTQCATSPHVNNRDRMTVTKSKNNQLFV	960
Db	901	VVNADYTISGTLPHSNGGNANSHYFTNP	SYHTLTQCATSPHVNNRDRMTVTKSKNNQLFV	960
Qy	961	NLKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN		1020
Db	961	NLKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN		1020
Qy	1021	CSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT		1080
Db	1021	CSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT		1080
Qy	1081	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE		1140
Db	1081	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE		1140

RESULT 3

ADE71305

ID ADE71305 standard; protein; 1192 AA.

XX

AC ADE71305;

XX

DT 29-JAN-2004 (first entry)

XX

DE Novel human protein #59.

XX

KW human; novel protein; drug.

XX

OS Homo sapiens.

XX

PN JP2002345493-A.

XX

PD 03-DEC-2002.

XX

PF 29-MAR-2001; 2002JP-00049046.

XX

PR 29-MAR-2001; 2001JP-00095524.

XX

PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
 XX
 DR WPI; 2003-460885/44.
 DR N-PSDB; ADE71243.
 XX
 PT A gene and a protein encoded by it, used in drugs.
 XX
 PS Disclosure; Page 242-247; 257pp; Japanese.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC human proteins. The DNA and protein sequences of the invention are used
 CC in drugs. The present amino acid sequence represents a novel human
 CC protein of the invention.
 XX
 SQ Sequence 1192 AA;

Query Match 100.0%; Score 6744; DB 7; Length 1192;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	53	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	112
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	113	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	172
Qy	121	PNTCQCEPGWG GTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Db	173	PNTCQCEPGWG GTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	232
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	240
Db	233	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	292
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCD AATGQCHCSP	300
Db	293	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCD AATGQCHCSP	352
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVS GACLCEAGFAGERCEARLCPEGLY	360
Db	353	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVS GACLCEAGFAGERCEARLCPEGLY	412
Qy	361	GIKCDKRCPC HLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQ QICSCQNGAD	420
Db	413	GIKCDKRCPC HLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQ QICSCQNGAD	472
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKND A VCS PVDG SCTCKAGWHGV	480
Db	473	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKND A VCS PVDG SCTCKAGWHGV	532
Qy	481	DCSIRCPSGTWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPQDGT YGLNCAE	540
Db	533	DCSIRCPSGTWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPQDGT YGLNCAE	592
Qy	541	RDCSHADGCHPTTGHCRC L PGWSGVHCD SVCAEGRWGPNC SLPCYCKNGAS CSPDDGIC	600

Db	593	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSDVCAEGRWGPNCSLPCYCKNGASCSPDDGIC	652
Qy	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVC	660
Db	653	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVC	712
Qy	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Db	713	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	772
Qy	721	GAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF	780
Db	773	GAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF	832
Qy	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIVGNLN	840
Db	833	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIVGNLN	892
Qy	841	SLSRTSTALPADSYQIGAIAGIIILVLVVLFLLLALFIIYRHKQKGKESSMPAVTYTPAMR	900
Db	893	SLSRTSTALPADSYQIGAIAGIIILVLVVLFLLLALFIIYRHKQKGKESSMPAVTYTPAMR	952
Qy	901	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFV	960
Db	953	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFV	1012
Qy	961	NLKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN	1020
Db	1013	NLKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN	1072
Qy	1021	CSLSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT	1080
Db	1073	CSLSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT	1132
Qy	1081	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1140
Db	1133	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1192

RESULT 4

AAE27986

ID AAE27986 standard; protein; 586 AA.

XX

AC AAE27986;

XX

DT 27-JAN-2003 (first entry)

XX

DE Human EGF-family protein #2.

XX

KW Human; EGF-family protein; novel human protein; NHP; drug discovery;
 KW restriction fragment length polymorphism analysis; forensic biology;
 KW toxicity; infectious disease; biological disorder; medical disorder;
 KW mental disorder; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200272611-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 06-MAR-2002; 2002WO-US007477.
 XX
 PR 12-MAR-2001; 2001US-0275013P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Yu X, Miranda M;
 XX
 DR WPI; 2002-723315/78.
 DR N-PSDB; AAD46319.
 XX
 PT New novel human nucleic acids useful for e.g. identifying protein coding
 PT sequences and mapping unique genes to a particular chromosome, as DNA
 PT markers for restriction fragment length polymorphism analysis, or in
 PT forensic biology.
 XX
 PS Claim 2; Page 40-42; 42pp; English.
 XX
 CC The present sequence is EGF-family protein, a novel human protein (NHP).
 CC The NHP sequences are useful for mapping unique genes to a particular
 CC chromosome; as DNA markers for restriction fragment length polymorphism
 CC analysis; in forensic biology; in defining and monitoring both drug
 CC action and toxicity; in identifying, selecting and validating novel
 CC molecular targets for drug discovery; in microarrays or other assay
 CC formats to screen collections of genetic material from patients who have
 CC a particular medical condition. The NHP peptides, fusion proteins,
 CC antibodies, antagonists and agonists can be used for detecting mutant
 CC NHPs or inappropriately expressed NHPs for the diagnosis of disease; for
 CC screening drugs for treatment of symptomatic or phenotypic manifestations
 CC of perturbing the normal function of NHP in the body and to treat
 CC diseases including infectious, mental, biological, or medical diseases or
 CC disorders. They are also used in gene therapy
 XX
 SQ Sequence 586 AA;

Query Match 53.4%; Score 3601; DB 5; Length 586;
 Best Local Similarity 100.0%; Pred. No. 4.8e-170;
 Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	1	MVISLNSCLSFICLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Qy	121	PNTQCCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Db	121	PNTQCCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Qy	181	RCEQGTYGNDCHQRCQCCQNGATCDHVTGECRCPPGYTGAFCEDLCPPGKHGPQCEQRCPC	240

Db 181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC 240
 Qy 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
 Qy 301 GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY 360
 Qy 361 GIKCDKRCPCHLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 361 GIKCDKRCPCHLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
 Qy 421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV 480
 Qy 481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE 540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE 540
 Qy 541 RCDCSHADGCHPTTGHCRCLPGWSGVHCDVCAEGRWGPNCSLPCY 586
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 541 RCDCSHADGCHPTTGHCRCLPGWSGVHCDVCAEGRWGPNCSLPCY 586

RESULT 5

ABG08033

ID ABG08033 standard; protein; 878 AA.

XX

AC ABG08033;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #8024.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS72220.

XX
 KW Membrane associated protein; secreted protein; human; mouse; rat;
 KW INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 295; TANGO 354;
 KW TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;
 KW haematopoietic disorder; neural disorder; hepatic disorder;
 KW neoplastic disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200100673-A1.
 XX
 PD 04-JAN-2001.
 XX
 PF 29-JUN-2000; 2000WO-US018198.
 XX
 PR 30-JUN-1999; 99US-00345464.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;
 XX
 DR WPI; 2001-050128/06.
 DR N-PSDB; AAF27787.
 XX
 PT Isolated secreted or transmembrane proteins are used for diagnosis and
 PT treatment of neoplastic and hematopoietic disorders e.g. T cell
 PT disorders, cancer and tumors.
 XX
 PS Claim 9; Page 227-229; 294pp; English.
 XX
 CC The present invention provides the protein and coding sequences for a
 CC number of membrane associated and secreted proteins from human, mouse and
 CC rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
 CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
 CC involved in signal transduction and the sequences can be used in the
 CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
 CC and haematopoietic disorders
 XX
 SQ Sequence 1050 AA;

Query Match 37.2%; Score 2506.5; DB 4; Length 1050;
 Best Local Similarity 40.5%; Pred. No. 8.4e-116;
 Matches 490; Conservative 111; Mismatches 345; Indels 263; Gaps 30;

Qy 14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNW----- 66
 ||| : | || ||| || |||:: | :||: || : | |
 Db 9 LLLAVGLRLAGTLNPSDNTCSFEWSEFTTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66
 Qy 67 -----FKCTRHRVSYR-----TAY 80
 | | : |
 Db 67 SPQTQRKLLASRDSFCMVCVGAGVQWRDRSALQPQTGNALSMRPQPRVLSGAPSLASPGH 126
 Qy 81 RHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGWGGTNCSSA-- 138
 || :|:: ||| |||| | ||| || :|||::||| ||| ||| | :|||
 Db 127 TVVVKTDHRQRLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCQCVPGWRGDDCSSAPN 186
 Qy 139 ----CDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQR 194

Db	187	CLQPCTPGYYGPACQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGT-----	237
Qy	195	CQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGEC	254
Db	238	-----SGFFC-----PSTH-----PCQNGGVFQTPQGS	262
Qy	255	CPSGWMGTVCGQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVG	314
Db	263	CPFGWMGTICSLPCPEGFHGPNCSEQECRCHNGGLCDRFTGQCRCAPGYTGDRCREEC	322
Qy	315	TYGVLCAETCQCVNNGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC	374
Db	323	RFGQDCAETCDCAPDARCFPANGACLCEHGF"TGDRCTDRLCPDGFYGLSCQAPCTCDREH	382
Qy	375	THSCHPMSEGCACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGF	434
Db	383	SLSCHPMNGECSCLPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGVCQATSGLCQCAPGY	442
Qy	435	KGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGVDCSIRCPSTWGF	494
Db	443	TGPHCASLCPDPTYGVNCSARCSCEAIACSPIDGECVCKEGWQRGNCSPVCPPTWGF	502
Qy	495	CNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTT	554
Db	503	CNASCQCAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGFGEGCASRCDCHSDGCDPVH	562
Qy	555	GHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTT	614
Db	563	GRCQCQAGWMGARCHLSCEPLWGVNCSNTCTCKNGGTCLPENGNCVCAPGFRG	622
Qy	615	CSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICT	674
Db	623	CQPGRYGKR-----CVP-----CK	636
Qy	675	CTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCT	734
Db	637	CANHSFCHPSNGTCYCLAGWTGPDSCQPCPPGHWGENCAQTCQCHHGGTCHPQDGS	696
Qy	735	PGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDDHISGQCTCRTGFMGRHCEQKCP	794
Db	697	LGWTGHHCLEGCPLGTFGANCSQPCQCGPGEKC-----HPE-----	732
Qy	795	GYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLRTSTALPAD--	852
Db	733	-----TGACVCPPGHSGAPCR-----IG----IQEPFTVMPPTPV	763
Qy	853	SY-QIGAIAGIIILVLVFLFLLALFIIYRHKQKGKESMPAVTYTPAMRVVNADYTISGT	911
Db	764	AYNSLGAVIGIAVLGSLVVALVALFIGYRHWQKGKEHHHLAVAYSSG-RLDGSEYVMPDV	822
Qy	912	LPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKN-VNPGKR	970
Db	823	PP-----SYSHYYSNPSYHTLSQCSNP PPPPNK-----VPGPLFASLQNP	866
Qy	971	GPVG-DCTGTLPADWKH-----GGYLNELGAFGLDRSYMGKSL-----KDLGKNSEY	1016

Db 867 GAQGHDNHTTLPADWKHRREPPPGPLDR-GSSRLDRSYSYSYNGPGPFYDKGLISEEEL 925

Qy 1017 NSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYE 1076
 :| || ||||| |||||:| | | | :||| | :| | |

Db 926 GASVASL-SSENPYATIRDLPSLPGGPRESSYMEMKGPPSGSAPRQPPQFWDSQRRR--- 981

Qy 1077 VEPTVSVVQGVFSNNGRL-----SQDP-----YDLPKNSHIPCHYDLLPVRDS 1119
 :| | : | || | || ||||| ||| |||

Db 982 -QPQPQRDSGTYEQPSPLIHDRDSVGSQPPLPPGLPPGHYDSPKNSHIPGHYDLPPVRHP 1040

Qy 1120 SSSP-KQED 1127
 | | ::|

Db 1041 PSPPLRRQD 1049

RESULT 7

AAG79417

ID AAG79417 standard; protein; 994 AA.

XX

AC AAG79417;

XX

DT 25-OCT-2002 (first entry)

XX

DE CADHP-6, Incyte ID No: 4097936CD1.

XX

KW Human; cell adhesion protein; CADHP; AIDS; Alzheimer's disease;

KW acquired immunodeficiency syndrome; thymic dysplasia; epilepsy;

KW renal tubular acidosis; congenital glaucoma; cancer; atherosclerosis;

KW Parkinson's disease.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	Domain	1. .609
FT		/label= Sushi_repeat
FT		/note= "Identified by BLAST-DOMO"
FT	Peptide	1. .29
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .28
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .25
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .24
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .22
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .20
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .20
FT		/label= Signal_cleavage
FT		/note= "Identified by SPSCAN"

FT	Peptide	1. .19
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .18
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Peptide	1. .16
FT		/label= Signal_peptide
FT		/note= "Identified by HMMER"
FT	Modified-site	30
FT		/note= "Potentially phosphorylated"
FT	Modified-site	38
FT		/note= "Potentially phosphorylated"
FT	Domain	101. .131
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	120. .131
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Domain	120. .131
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Binding-site	127. .129
FT		/label= Cell_attachmnt_sequence
FT		/note= "Identified by MOTIFS"
FT	Peptide	133. .161
FT		/label= Type_III_EGF-like_signature
FT		/note= "Identified by BLIMPS-PRINTS"
FT	Domain	138. .576
FT		/label= Sushi_repeat
FT		/note= "Identified by BLAST-DOMO"
FT	Domain	144. .174
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Modified-site	152
FT		/note= "Potentially glycosylated"
FT	Modified-site	153
FT		/note= "Potentially glycosylated"
FT	Modified-site	154
FT		/note= "Potentially phosphorylated"
FT	Domain	187. .216
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	205. .216
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	229. .259
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	248. .259
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	248. .259
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Domain	248. .252
FT		/label= Sushi_domain_protein

FT		/note= "Identified by BLIMPS-PFAM"
FT	Modified-site	271
FT		/note= "Potentially glycosylated"
FT	Domain	272. .302
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Peptide	284. .302
FT		/label= Type_III_EGF-like_signature
FT		/note= "Identified by BLIMPS-PRINTS"
FT	Domain	291. .302
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	291. .302
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Domain	315. .345
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	334. .345
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	334. .345
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	346
FT		/note= "Potentially phosphorylated"
FT	Modified-site	355
FT		/note= "Potentially phosphorylated"
FT	Domain	365. .391
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	380. .391
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	380. .391
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	392
FT		/note= "Potentially glycosylated"
FT	Domain	404. .434
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Domain	423. .434
FT		/label= EGF-like_domain_signature_1
FT		/note= "Identified by MOTIFS"
FT	Domain	423. .434
FT		/label= EGF-like_domain_signature_2
FT		/note= "Identified by MOTIFS"
FT	Modified-site	446
FT		/note= "Potentially glycosylated"
FT	Domain	447. .477
FT		/label= EGF-like_domain
FT		/note= "Identified by HMMER-PFAM"
FT	Modified-site	448
FT		/note= "Potentially phosphorylated"
FT	Modified-site	460
FT		/note= "Potentially phosphorylated"

```

FT      Domain      466. .477
FT                               /label= EGF-like_domain_signature_2
FT                               /note= "Identified by MOTIFS"
FT      Modified-site 476
FT                               /note= "Potentially glycosylated"
FT      Domain      490. .520
FT                               /label= EGF-like_domain
FT                               /note= "Identified by HMMER-PFAM"
FT      Modified-site 491
FT                               /note= "Potentially glycosylated"
FT      Domain      509. .520
FT                               /label= EGF-like_domain_signature_1
FT                               /note= "Identified by MOTIFS"
FT      Domain      509. .520
FT                               /label= EGF-like_domain_signature_2
FT                               /note= "Identified by MOTIFS"
FT      Domain      533. .563
FT                               /label= EGF-like_domain
FT                               /note= "Identified by HMMER-PFAM"
FT      Modified-site 535
FT                               /note= "Potentially phosphorylated"
FT      Domain      552. .563
FT                               /label= EGF-like_domain_signature_1
FT                               /note= "Identified by MOTIFS"
FT      Domain      552. .563
FT                               /label= EGF-like_domain_signature_2
FT                               /note= "Identified by MOTIFS"
FT      Modified-site 566
FT                               /note= "Potentially phosphorylated"
FT      Modified-site 575
FT                               /note= "Potentially glycosylated"
FT      Domain      576. .606
FT                               /label= EGF-like_domain
FT                               /note= "Identified by HMMER-PFAM"
FT      Modified-site 581
FT                               /note= "Potentially phosphorylated"
FT      Domain      595. .606
FT                               /label= EGF-like_domain_signature_1
FT                               /note= "Identified by MOTIFS"
FT      Domain      595. .606
FT                               /label= EGF-like_domain_signature_2
FT                               /note= "Identified by MOTIFS"
FT      Domain      603. .614
FT                               /label= Sushi_domain_protein
FT                               /note= "Identified by BLIMPS-PFAM"
FT      Domain      619. .648

```

```

Query Match          36.8%;  Score 2482.5;  DB 5;  Length 994;
Best Local Similarity 41.6%;  Pred. No. 1.2e-114;
Matches 482;  Conservative 110;  Mismatches 347;  Indels 221;  Gaps 27;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : | || ||| |||:: | :||: || : | | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFEWESFTTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || | :|:: ||| |||| | || || :|||::||| ||| |||

```

Db	67	QPTVVYRTVYRQVVKTDHRQRLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCQCVPGW	126
Qy	131	GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND	190
Db	127	: : : : : :	
Db	127	RGDDCSSECAPGMWGPQCDKPCSCGNSSCDPKSGVCSGPSGLQPPNCLQPCTPGYYGPA	186
Qy	191	CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPPGKHGHPQCEQRCPCQNGGVCHHVT	250
Db	187	: :	
Db	187	CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHPCQNGGVFQTPQ	245
Qy	251	GECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDE	310
Db	246	: : : : ::	
Db	246	GSCSCPPGWMGTICSLPCPEGFHGPNCSEQECRCHNGGLCDRFTGQCRCAPGYTGDRCREE	305
Qy	311	CPVGTGYVLCAETCQCVNNGGKCYHVSAGACLCEAGFAGERCEARLCPGELYGIKCDKRCPC	370
Db	306	: : : : : :	
Db	306	CPVGRFGQDCAETCDCAPDARCFPANGACLCEHGFTGDRCTDRLCPDGFYGLSCQAPCTC	365
Qy	371	HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC	430
Db	366	:: : :: : :: :: : : : : :	
Db	366	DREHSLSCHPMNGECSCLPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGVCQATSGLCQC	425
Qy	431	APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGVDCSIRCPSGT	490
Db	426	: : : :: : : : :	
Db	426	APGYTGPHCASLCPDPTYGVNCSARCSCENAIACSPIDGECVCKEGWQRGNCSVPCPPGT	485
Qy	491	WGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTGYGLNCAERCDCSHADGC	550
Db	486	: : : : :	
Db	486	WGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASRCDCHSDGC	545
Qy	551	HPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTT	610
Db	546	: : : :	
Db	546	DPVHGRCQCQAGWMGARCHLSCPEGLWGVNCSNTCTCKNGGTCLPENGNCVCAPGFRGPS	605
Qy	611	CQRICSPGFYGHRCSTCPQCQVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA	670
Db	606	:	
Db	606	CQRSCQPGRYGKR-----CVP-----	621
Qy	671	GICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGE	730
Db	622	-----CKCANHSFCHPSNGT	636
Qy	731	CKCTPGWTGLYCTQRCPLGFYKGKCALICQCQNGADCDDHISGQCTCRTGFMGRHCEQKCP	790
Db	637	: : ::	
Db	637	CYCLAGWTGPDCSQRCPLGTFGANCSQPCQCQGPGEKC-----HPE-----	676
Qy	791	SGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTALP	850
Db	677	: : :	
Db	677	-----TGACVCPPGHSGAPCR-----IG-----IQEPFTVMP	703
Qy	851	AD--SY-QIGAIAGIIILVLVLFLLALFIIYRHKQKGKESSMPAVTYTPAMRVVNADYT	907
Db	704	: : : : :: : : : :	
Db	704	TTPVAYNSLGAVIGIAVLGSLVVALVALFIGYRHWQKGKEHHHLAVAYSSG-RLDGSEYV	762
Qy	908	ISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVN-KNVN	966
Db	763	: : :: : :: : : :	
Db	763	MPDVPP-----SYSHYYSNPSYHTLSQCSPPPPPNK-----VPGPLFASLQKPER	808

Qy 967 PGKRGVPG-DCTGTLPADWKH-----GGYLNEIGAFGLDRSYM GKSL-----KDLGK 1012
 || | | | | | | | | | | : | : | | | | | |
 Db 809 PG--GAQGHDNHTTLPADWKHRREPPPGPLDR-GSSRLDRSYSYSYSNGPGPFYNKGLIS 865
 Qy 1013 NSEYNSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSP-----ARR 1057
 | : | | | | | | | : | | | | : | | | | |
 Db 866 EEELGASVASL-SSENPYATIRDLPSPGGPRESSYMEMKGPPSGSPPRQPPQFWDSQRR 924
 Qy 1058 DSPYAEINNSTSANRNVYEVEPTVSVVQGVFSNNGRLSQDP-----YDLPKNSHIP 1108
 | : : | | : : : : : | | | | | | | | | |
 Db 925 RQPQQRDSGT-----YE-QPSPL---IHDRDSVGSQPPLPPGLPPGHYDSPKNSHIP 973
 Qy 1109 CHYDLLPVRDSSSSP-KQED 1127
 |||| || | | : : |
 Db 974 GHYDLPPVRHPPSPPLRRQD 993

RESULT 8

AAB66269

ID AAB66269 standard; protein; 636 AA.

XX

AC AAB66269;

XX

DT 05-APR-2001 (first entry)

XX

DE Rat TANGO 272 SEQ ID NO: 20.

XX

KW Membrane associated protein; secreted protein; human; mouse; rat;
 KW INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 295; TANGO 354;
 KW TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;
 KW haematopoietic disorder; neural disorder; hepatic disorder;
 KW neoplastic disease.

XX

OS Rattus sp.

XX

PN WO200100673-A1.

XX

PD 04-JAN-2001.

XX

PF 29-JUN-2000; 2000WO-US018198.

XX

PR 30-JUN-1999; 99US-00345464.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;

XX

DR WPI; 2001-050128/06.

DR N-PSDB; AAF27791.

XX

PT Isolated secreted or transmembrane proteins are used for diagnosis and
 PT treatment of neoplastic and hematopoietic disorders e.g. T cell
 PT disorders, cancer and tumors.

XX

PS Claim 9; Page 238-240; 294pp; English.

XX

CC The present invention provides the protein and coding sequences for a
CC number of membrane associated and secreted proteins from human, mouse and
CC rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
CC involved in signal transduction and the sequences can be used in the
CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
CC and haematopoietic disorders

XX

SQ Sequence 636 AA;

Query Match 28.3%; Score 1909; DB 4; Length 636;

Best Local Similarity 45.1%; Pred. No. 1.6e-86;

Matches 328; Conservative 77; Mismatches 212; Indels 110; Gaps 9;

Qy 260 MGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVL 319

|| :| ||||| | ||:| |:| |||| | | ||||| |:| | |:| :| ||||| :|

Db 1 MGVICSLPCPEGFHGPNECTQECRCHNGGLCDRETGOCHCAPGYIGDRCREECPVGRFGQD 60

Qv 320 CAETCQCVNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHLENTHSCH 379

[illegible]

Db 61 CAETCDCAPGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQDPCTCDPEHSLSCH 120

Qv 380 PMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDC 439

|| || : | : | || : | : | : | : | : | : | : | : |

Db 121 PMHGECSQPGWAGLHCNESCPODTHGAGCQEHCLCLHGGVCLADSGLCRCAPGYTGPHC 180

Qy 440 STPCPLGTYGINCSSRCGCKNDVCS PVDGSCTCKAGWHGVDCSIRCPSTWGFGCNLTC 499

$$: \quad || \quad ||||| \quad | \quad :| \quad |||||:| \quad || \quad || \quad :||: \quad || \quad ||||| \quad || \quad :|$$

Db 181 ANLCPPNTYGINCSSHCSCENAIACSPVDGTCICKEGWQRGNCSVPCPPGTWGFSCNASC 240

Qy 500 QCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGHCR 559

|| : | |:) ||| |||| | :||| | :|)| ||| |:}||| | |||||

Db 241 QCAHEGVCSPTGACTCTPGWRGVHCQLPCPKGQFGEGCASVCDCHSDGCDPVHGHCR 300

Qy 560 LPGWSGVHCD SVCAEGRWGPNC SLPCYCKNGASC SPDDGICECAPGFRGTTCQRICSPGF 619

|| | | || || || | |||| :| |:| | ||||| :| || | ||

Db 301 QAGWMGTRCHLPCPEGFWGANCSNACTCKNGGTCVPENGNCVCAPGFRGPSCQRPCPPGR 360

Qy 620 YGHRCSQTC PQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNG 679

[illegible]

Db 361 YGKR-----CVP-----CKCNNHS 374

Qy 680 TCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTG 739

:|:| | :| | || | |||: ||| ||| | | |:| | | | |

Db 375 SCHPSDGTCSCLAGWTGPDCSESCPPGHWGLKCSQPCQCHHGATCHPQDGSCVCIPGWTG 434

Qy 740 LYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKCPSGTYGYGCR 799

[illegible]

Db 435 PNCSEGCP SRMFGVNC SQLCQCDPGEMC-----HPE----- 465

Qy 800 QICDCLNNSTCDHITGTCYCS PGWK GARC DQAGVIIVGNLNSLSRTSTALPADSYQIGAI 859

$$|| \quad | \quad | \quad || \quad || \quad | \qquad || : \quad | : \quad | : | \qquad : || :$$

Db 466 -----TGACVCPPGHSGAHCK-----VGSQESFTIMPTS-PVIHNSLGAV 504

Qy 860 AGIIILVLVFLFLLALFIIYRHKQKGKESMPAVTYTPAMRVVNADYTISGTLPHSNGGN 919

|| : | : | : | : || || | || | || || | || || | || : | : : | | : |

Db 505 IGI AVL GTL V VAL VAL F I G Y R H W Q K G K E H E H L A V A Y S T G - R L D G S D Y V M P D V S P - - - - - 557

Qy 920 ANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKNVNPGRGPVGDCTGT 979
 : |||::|||||:|: :| | :|||: : | |
 Db 558 SYSHYSNPSYHTLSQCSFNPPPPN-----KIPGSQLFVSSQASERPNRNHGRDNHAT 610

Qy 980 LPADWKH 986
 |||||
 Db 611 LPADWKH 617

RESULT 9

ADA21141

ID ADA21141 standard; protein; 1350 AA.

XX

AC ADA21141;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human secreted protein SECP-46 SEQ ID NO:46.

XX

KW human; secreted protein; SECP; anti-HIV; antiallergic; antiinflammatory;

KW antianaemic; antiparkinsonian; nootropic; anticonvulsant;

KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;

KW cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;

KW antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic;

KW antiparasitic; antihelminthic; antipsoriatic; uropathic;

KW ophthalmological; antirheumatic; haemostatic; antibacterial; virucide;

KW protozoacide; fungicide; gene therapy; cell proliferative disorder;

KW arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;

KW paroxysmal nocturnal haemoglobinuria; polycythaemia vera; psoriasis;

KW primary thrombocytopaenia; cancer; developmental disorder;

KW renal tubular acidosis; anaemia; mental retardation;

KW neurological disorder; Alzheimer's disease; Parkinson's disease;

KW epilepsy; autoimmune disorder; inflammatory disorder; AIDS; allergy;

KW asthma; autoimmune thyroiditis; contact dermatitis; Crohn's disease;

KW diabetes mellitus; glomerulonephritis; Goodpasture's syndrome; gout;

KW Graves' disease; Hashimoto's thyroiditis; irritable bowel syndrome;

KW multiple sclerosis; osteoarthritis; osteoporosis; pancreatitis;

KW Reiter's syndrome; rheumatoid arthritis; Sjogren's syndrome; uveitis;

KW infection.

XX

OS Homo sapiens.

XX

PN WO2003068943-A2.

XX

PD 21-AUG-2003.

XX

PF 13-FEB-2003; 2003WO-US004712.

XX

PR 13-FEB-2002; 2002US-0357002P.

PR 06-MAR-2002; 2002US-0362439P.

PR 19-MAR-2002; 2002US-0366041P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Lehr-Mason PM, Kable AE, Elliott VS, Marquis JP, Baughn MR;

PI Chawla NK, Tran UK, Jin P, Tang YT, Zebarjadian Y, Swarnakar A;

PI Hafalia AJA, Cocks BG, Warren BA, Emerling BM, Pearson CI, Chien D;
PI Peterson DP, Fu GK, Yue H, Jackson AA, Jiang X, Hawkins PR, Lal PG;
PI Khare R, Lee S, Lee SY, Richardson TW, Chang H;

XX

DR WPI; 2003-689669/65.

DR N-PSDB; ADA21192.

XX

PT New human secreted proteins and polynucleotides, useful for diagnosing,
PT treating or preventing autoimmune or inflammatory disorders (e.g. AIDS,
PT allergy, asthma or anemia), multiple sclerosis, osteoporosis, cancer or
PT hepatitis.

XX

PS Claim 1; Page 249-252; 295pp; English.

XX

The present sequence represents a human secreted protein (I) designated SECP-46. (I) have anti-HIV, antiallergic, antiinflammatory, antianaemic, antiparkinsonian, nootropic, anticonvulsant, antiarteriosclerotic, antiasthmatic, immunosuppressive, antithyroid, cytostatic, hepatotropic, dermatological, antidiabetic, nephrotropic, antigout, thyromimetic, neuroprotective, osteopathic, antiarthritic, antiparasitic, antihelminthic, antipsoriatic, uropathic, ophthalmological, antirheumatic, haemostatic, antibacterial, virucide, protozoacide and fungicide activities, and can be used in gene therapy. The human secreted proteins (SECP), polynucleotides, agonists and antagonists of the present invention are useful for diagnosing, treating or preventing disorders associated with aberrant expression of SECP, particularly cell proliferative disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary thrombocytopaenia or cancer), developmental disorders (e.g. renal tubular acidosis, anaemia or mental retardation), neurological disorders (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), autoimmune/inflammatory disorders (e.g. AIDS, allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral, bacterial, fungal, parasitic, protozoan or helminthic infections. The SECP and polynucleotides are also useful in assessing the effects of exogenous compounds on the expression of nucleic acids secreted proteins. The polynucleotides encoding SECP are useful for creating transgenic animals to model human disease.

XX

SQ Sequence 1350 AA;

Query Match 28.1%; Score 1897; DB 6; Length 1350;
Best Local Similarity 40.4%; Pred. No. 1.3e-85;
Matches 334; Conservative 90; Mismatches 325; Indels 78; Gaps 15;

```

QY      94  CCPGFYESGEMC-----VPHCADKCVHGRCIAPNT-----CQCEPGWGGTNC 135
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      523  CDPGLY--GRFCHLTCPFWAFGPGCSEEC---QCVQPHTQSCDKRDGSCSCKAGFRGERC 577

QY      136  SSACDGDHWGPHCTSRCQCKNGALCNPITGAC--HCAAGFRGWRCEDRCEQGTYNDC HQ 193
      : | : : || | | | | | | | | | | | | | | | | | | | | |
Db      578  OAECELGYFGPGCWQACTCPVGVACDSVSGECGRCPAGFQGEDCGQECVPVGTFGVNCSS 637

```

Qy	194	RCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPFGKHGPQCEQRCPC-QCNGGVCHHVTGE	252
Db	638	SCSC-GGAPCHGVTGQCRCPPGRTGEDCEADCPEGRWGLGCQEICPACQHAARCDPETGA	696
Qy	253	CSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECP	312
Db	697	CLCLPGFVGSRCQDVCAPAGWYGPSCQTRCSCANDGHCHPATGHCSCAPGWTGFSCQACD	756
Qy	313	VGTYGVLCATCQCQVNG-GKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC	371
Db	757	TGHWGPDCSHPCNCSAGHGSCDAISGLCLCEAGYVGRPRCEQQ-CPQGHFGPGCEQLCQC-	814
Qy	372	LENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCA	431
Db	815	-QHGAACDHVSGACTCPAGWRGTFCEHACPAFFGLDCRSACNCTAGAACDAVNGSCLCP	873
Qy	432	PGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSGCTCKAGWHGVDCSIRCPSGTW	491
Db	874	AGRRGPRCAETCPAHTYGHNC SQACACFNGASCDPVHGQCHCAPGWMGPSCLECLPRDV	933
Qy	492	GFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCH	551
Db	934	RAGCRHSGGCLNGGLCDPHTGRCLCPAGWTGDKCQSPCLRGWFGEACAQRCS CPPGAACH	993
Qy	552	PTTGHCRCPLPGWSGVHCDSVCAEGRWGPNC SLPCYCK-NGASCSPDDGICECAPGFRGTT	610
Db	994	HVTGACRCPPGFTGSGCEQACPPGSFGEDCAQMCQCPGENPACHPATGTCTSCAAGYHG	1053
Qy	611	CQRICSPGFYGHRC SQTCPCQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA	670
Db	1054	CQQRCPPGRYGPGCEQLC-GCL-NGGSCDAATGACRCPTGFLGTDCNLTC PQGRFGPNCT	1111
Qy	671	GICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGE	730
Db	1112	HVCGCGQGAACDPVTGTCLCPPGRAGVRCERGC PQNRFGVGCEHTCSCRNGGLCHASNGS	1171
Qy	731	CKCTPGWTGLYCTQRCPLGFYGKDCAL-----	757
Db	1172	CSCGLGWTGRHCELACPPGRYGAACHLECSCHNNSTCEPATGTCTRCGPGFYQACEHPCP	1231
Qy	758	-----ICQCQNGADC DHISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNN	807
Db	1232	PGFHGAGCQGLCWCQH GAPCDPISGRCLCPAGFHGHFCERGCEPGSFGEGBCHQRCD CDGG	1291
Qy	808	STCDHITGTCYCS PGWKGARCDQAGVIVGNLNSLSRTST-ALPADS	853
Db	1292	APCDPVTGLCLCPPGRSGATCNLGGPLRLPENPSLAQGSAGTLPASS	1338

RESULT 10

ABJ37904

ID ABJ37904 standard; protein; 1577 AA.

XX

AC ABJ37904;

XX

DT 22-MAY-2003 (first entry)

XX

DE NOVX protein sequence SEQ ID No 54.
 XX
 KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
 KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
 KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
 KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
 KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
 KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
 KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
 KW immunogen; non-human transgenic animal; gene therapy.
 XX
 OS Unidentified.
 XX
 PN WO200281517-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 22-JAN-2002; 2002WO-US002064.
 XX
 PR 19-JAN-2001; 2001US-0262892P.
 PR 23-JAN-2001; 2001US-0263598P.
 PR 24-JAN-2001; 2001US-0263799P.
 PR 25-JAN-2001; 2001US-0264117P.
 PR 25-JAN-2001; 2001US-0264139P.
 PR 26-JAN-2001; 2001US-0264478P.
 PR 30-JAN-2001; 2001US-0263351P.
 PR 02-MAR-2001; 2001US-0272870P.
 PR 14-MAR-2001; 2001US-0275927P.
 PR 14-MAR-2001; 2001US-0275990P.
 PR 15-MAR-2001; 2001US-0276449P.
 PR 20-MAR-2001; 2001US-0277358P.
 PR 23-MAR-2001; 2001US-0278151P.
 PR 29-MAR-2001; 2001US-0279857P.
 PR 20-APR-2001; 2001US-0285140P.
 PR 20-APR-2001; 2001US-0285141P.
 PR 30-APR-2001; 2001US-0287484P.
 PR 17-MAY-2001; 2001US-0291701P.
 PR 08-JUN-2001; 2001US-0296960P.
 PR 10-JUL-2001; 2001US-0304353P.
 PR 10-JUL-2001; 2001US-0304355P.
 PR 12-JUL-2001; 2001US-0304886P.
 PR 09-AUG-2001; 2001US-0311289P.
 PR 13-AUG-2001; 2001US-0311975P.
 PR 16-AUG-2001; 2001US-0312937P.
 PR 18-OCT-2001; 2001US-0330227P.
 PR 29-NOV-2001; 2001US-0334198P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Decristofaro MF, Padigar M, Miller C, Tchernev V, Zhong H;
 PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
 PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
 PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
 PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;

PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;
XX
DR WPI; 2003-058504/05.
DR N-PSDB; ABT33369.

XX
PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.

XX
PS Claim 1; Page 135-136; 672pp; English.

XX
CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
CC infections. They can also treat disorders such as e.g., Alzheimer's
CC disease or a stroke. The NOVX encoding nucleic acids are useful for
CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
CC for identifying a cell or tissue type in a biological sample, to amplify
CC DNA sequences from very small biological samples such as tissues e.g.
CC hair or skin or body fluids in forensic biology and as primers and probes
CC for use in identifying and/or cloning NOVX homologues in other cell
CC types. The NOVX proteins are useful as an immunogen to generate
CC antibodies which are useful for diagnostically monitoring protein levels
CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
CC useful for producing non-human transgenic animals which are useful for
CC studying the function and/or activity of NOVX protein and for identifying
CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
CC acids can be used in gene therapy. This sequence represents a NOVX
CC protein of the invention

XX
SQ Sequence 1577 AA;

Query Match 27.9%; Score 1879; DB 6; Length 1577;
Best Local Similarity 40.4%; Pred. No. 1.1e-84;
Matches 325; Conservative 83; Mismatches 318; Indels 78; Gaps 15;

QY	94	CCPGFYESGEMC-----VPHCADKCVHGRCIAPNT-----CQCEPGWGGTNC	135
Db	634	CDPGLY--GRFCHLACPPWAFGPGCSEEC---QCVQPHTQSCDKRDGSCSCKAGFRGERC	688
QY	136	SSACDGDHWGPHCTSRCQCKNGALCNPTTGAC--HCAAGFRGWRCEDRCEQGTYNDCHQ	193
Db	689	QAECEPGYFGPGCWQACTCPVGVACDSVSGECGKRCPAGFQGEDCGQECVPVGTFGVNCSS	748
QY	194	RCQCQNGATCDHVTGECRCPPGYTGAFCE-----DLCP-----	226
Db	749	SCSC--GGAPCHGVTGQCRCPPGRTGEDCEAGECEGLWGLGCQEICPACHNAARCDPETGA	807
QY	227	-----PGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCGQPCPE	270
Db	808	CLCLPGFVGSRCQDCEAGWYGPSQTMCS CANDGHCHQDTHGCSCAPGWTGFSCQRACDT	867
QY	271	GRFGKNCSQECQCHNG-GTCDAATGQCHCSPGYTGERC-QDECPVGTYGVLCAETCQCVN	328
Db	868	GHWGPDCSHPCNCSAGHGS CDAISGLCLCEAGYVGPRCEQSECPQHGFPGCEQRCQ CQH	927
QY	329	GGKCYHVS GACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHLENTHSCHPMSGECACK	388
Db	928	GAACDHVSGACTCPAGWRGTFCE-HACPAGFFGLDCRSACNC--TAGAACDAVNGSCLCP	984
QY	389	PGWSGLYCNETCSPGF-YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGT	447
Db	985	AGRRGPRCAESACPAHTYGHNC SQACACFNGASCDPVH GQCHCAPGWMGPSCLQACPAGL	1044
QY	448	YGINCSSRCGCKNDVAVCS PVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGGAC	507
Db	1045	YGDNCRHSCLCQNGGTCDPVSGHCACPEGWAGLACEVECLPRDVRAGCRHS GGCLNGGLC	1104
QY	508	NTLDGTCTCAPGWRGEKCELP--CQDGT YGLNCAERCDCSHADGCHPTTGHCRC L PGWSG	565
Db	1105	DPHTGRCLCPAGWTGDKCQSPAACAKGTFGPHCEGRACACRWGGPCHLATGACLCPPGWRG	1164
QY	566	VHCDSVCAEGRWGPNC SLPCYCKNGASCSPDDGICECAPGFRGTTQRICSPGFYGHRC S	625
Db	1165	PHLSAACLRGWFG EACAQRCS CPPGAACHHVTGACRCPPGFTGSGCEQACPPGSFGEDCA	1224
QY	626	QTC PQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPID	685
Db	1225	QMC-QCPGENPACHPATGTCTSCAAGYHGPSCQQRCP PGRYGP GCEQLCGCLNGGSCDAAT	1283
QY	686	RSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQR	745
Db	1284	GACRCPTGFLGTDCNLTC PQGRFGPNCTHVC GCGQGAACDPVTGTCLCPPGRAGVRCERG	1343
QY	746	CPLGFYGKDCALICQCQNGADCDHISGQCTCRTCRTGFMGRHCEQKCPSGTYGYGCRQICDCL	805
Db	1344	CPQNRFGVGCEHTCSCRNGGLCHASNGSCSGLGWTGRHCELACPPGRYGAACHLECSCH	1403
QY	806	NNSTCDHITGTCYCS PGWKGARCD	829
Db	1404	NNSTCEPATGTCTRCGPGFYGQACE	1427

RESULT 11

ADD78227

ID ADD78227 standard; protein; 1261 AA.

XX

AC ADD78227;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human CGDD-8.

XX

KW Anabolic; Hypertensive; Respiratory; Anti-HIV; Antiallergic;
 KW Neuroprotective; Nootropic; Antianemic; Antiarteriosclerotic;
 KW Antiinflammatory; Ophthalmological; Muscular; Hepatotropic;
 KW Neuroprotective; Antiasthmatic; Anticonvulsant; Virucide; Antibacterial;
 KW Fungicide; Antiparasitic; Protozoacide; Antihelminthic; Cytostatic;
 KW Cerebroprotective; Antiparkinsonian; Antipsoriatic; Antigout;
 KW Antidiabetic; Antiarthritic; Antirheumatic; Osteopathic; Gene therapy;
 KW human; cell growth; cell differentiation; cell death; CGDD;
 KW cell proliferative disorder; cancer; developmental disorder;
 KW neurological disorder; autoimmune disorder; inflammatory disorder;
 KW infection; reproductive disorder.

XX

OS Homo sapiens.

XX

PN WO2003077875-A2.

XX

PD 25-SEP-2003.

XX

PF 14-MAR-2003; 2003WO-US008310.

XX

PR 15-MAR-2002; 2002US-0364494P.

PR 29-MAR-2002; 2002US-0369129P.

PR 12-APR-2002; 2002US-0372511P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Kable AE, Tran UK, Hafalia AJA, Burford N, Honchell CD;
 PI Lehr-Mason PM, Duggan BM, Ramkumar J, Griffin JA, Richardson TW;
 PI Elliott VS, Jiang X, Jackson AA, Marquis JP, Chawla NK, Khare R;
 PI Becha SD, Lee SY, Swarnakar A, Yue H, Warren BA, Baughn MR, Lal PG;
 PI Lee S, Ho A, Gandhi AR, Yao MG;

XX

DR WPI; 2003-779081/73.

DR N-PSDB; ADD78266.

XX

PT New polypeptides and polynucleotides associated with cell growth,
 PT differentiation and death, useful for diagnosing, treating or preventing
 PT e.g. developmental, neurological, autoimmune, inflammatory or
 PT reproductive disorders.

XX

PS Claim 1; SEQ ID NO 8; 320pp; English.

XX

CC The present invention relates to novel human proteins (I; ADD78220-
 CC ADD78258) and their coding sequences (II; ADD78259-ADD78297), which are
 CC associated with cell growth, differentiation and death, referred to as
 CC CGDD-n proteins, where n is a number from 1 to 39. The CGDD proteins and
 CC their coding sequences are useful for diagnosing, treating or preventing

CC cell proliferative disorders (e.g. cirrhosis, hepatitis,
 CC arteriosclerosis, psoriasis, primary thrombocytopenia) or cancers (e.g.
 CC adenocarcinoma, sarcoma or cancers of the bone, bone marrow, brain,
 CC breast, colon, kidney, liver, lung or uterus), developmental disorders
 CC (e.g. renal tubular acidosis, Becker muscular dystrophy, gonadal
 CC dysgenesis, hypothyroidism or seizures), neurological disorders (e.g.
 CC Pick's disease, cataract, epilepsy, ischemic cerebrovascular disease,
 CC stroke, Alzheimer's disease, Parkinson's disease or dementia),
 CC autoimmune/inflammatory disorders (e.g. AIDS, allergies, anemia, asthma,
 CC diabetes mellitus, bronchitis, osteoporosis, osteoarthritis, rheumatoid
 CC arthritis, contact dermatitis or gout), viral, bacterial, fungal,
 CC parasitic, protozoan or helminthic infections, reproductive disorders
 CC (e.g. infertility, ectopic pregnancy, premature ovarian failure, delayed
 CC puberty or prostatitis) or disorders of the placenta (e.g. preeclampsia,
 CC choriocarcinoma, placenta previa, placental or maternal floor infarction
 CC or chronic villitis).

XX

SQ Sequence 1261 AA;

Query Match 27.8%; Score 1874.5; DB 7; Length 1261;
 Best Local Similarity 41.3%; Pred. No. 1.5e-84;
 Matches 328; Conservative 66; Mismatches 344; Indels 57; Gaps 10;

Qy 83 GEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRC-IAPNTCQCEPGWGGTNCSSACDG 141
 || : | | : | | | | : : | | | | : |
 Db 396 GEHTL-TEKFVCLDDSF--GHDCSLTCDDCRNGGTCLLGLDGCDCPEGWTGLICNETCPP 452

Qy 142 DHWGPHTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCONGA 201
 | : | : | | | : | | | : | | | : | | | : | | |
 Db 453 DTFGKNCSFSCSCQNGGTCDSVTGACRCPPGVSGTNCEDGCPKGYGKHCRKKCNCANRG 512

Qy 202 TCDHVTGECRCPPGYTGAFCEDLCP----- 227
 | : | | | | | | | | | |
 Db 513 RCHRLYGACLCDPGLYGRFCHLTCPWAFGPGCSEECQCVQPHQTQSCDKRDGSCSCKAGF 572

Qy 228 -----GKHGPQCEQRCPCQNGGVCHHVTGEC--CPSGWMGTVCQPCPEGRFG 274
 | | | | | | | | | : | | | | : | | | | | |
 Db 573 RGERCQAECELGYFGPGCWQACTCPVGVACDSVSGECGKRCPAGFQGEDCGQCEPVGTFG 632

Qy 275 KNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYH 334
 | | | | | | | | | | | | : : | | : | | : | | | |
 Db 633 VNCSSSCSC-GGAPCHGVTGQCRCPGRTGEDCEADCPEGHFGPGCEQRCQCHGAACDH 691

Qy 335 VSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCHPMMSGECACKPGWSGL 394
 | | | | | | : | | | | | : | | : : | | | | | |
 Db 692 VSGACTCPAGWRGTFCE-HACPAGFFGLDCRSACNC--TAGAACDAVNGSCLCPAGRRGP 748

Qy 395 YCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSS 454
 | | | | | | | | | : | | | | | : | | | | | | | |
 Db 749 RCAETCPAHTYGHNCSQACACFNGASCDFVHGQCHCAPGWMGPSCLOACPAGLYGDNCRH 808

Qy 455 RCGCKNDAVCSPVDGSCCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTC 514
 | : | | | | | | | : | | | : | | | | | : | |
 Db 809 SCLCQNGGTCDPVSGHCACPEGWAGLACEKECLPRDVRAGCRHSGGCLNGGLCDPHTGRC 868

Qy 515 TCAPGWRGEKCELPCQDGTYGLNCAERCDCSHADGCHPTTGHCRCPLGWSGVHCDVCAE 574
 | | | : | | | : | | : | | | | | | : | | : |

Db 869 LCPAGWTGDKCQSPCLRGWFGACAQRCSPPGAACHVHTGACRCPPGFTGSGCEQGCPP 928
 Qy 575 GRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHS 634
 ||:|| | | | | | | | | | | | | | | :| :| | |
 Db 929 GRYGPGCEQLCGCLNGGSCDAATGACRCPTGFLGTDCNLTCPQGRFGPNCTHVC-GCGQG 987
 Qy 635 SGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGW 694
 : | :|| | | | | | | | | | | | | | :| :| :| | | | |
 Db 988 AA-CDPVTGTCLCPPGRAGVRCERGCPQNRFGVGCEHTCSCRNGGLCHASNGSCSCGLGW 1046
 Qy 695 IGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKD 754
 | | | | | | | | :| | | | | :| | | | | | | | | | :|
 Db 1047 TGRHCELACPPGRYGAACHLECSCHNNSTCEPATGTCRCGPGFYGQACEHPCPPGFHGAG 1106
 Qy 755 CALICQCQNGADCDHISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHIT 814
 | :| ||:|| | | ||:|| | | | | ||: | | :|| | | | | | :| | :|
 Db 1107 CQGLCWCQHGA PCDPISGRCLCPAGFHGHFCERGCEPGSFGECHQRCDCDGGAPCDPVT 1166
 Qy 815 GTCYCSPGWKGARCD 829
 | | | | | | | :
 Db 1167 GLCLCPPGRSGATCN 1181

RESULT 12

ABP75770

ID ABP75770 standard; protein; 349 AA.

XX

AC ABP75770;

XX

DT 10-FEB-2003 (first entry)

XX

DE Human secretory polypeptide SPTM SEQ ID NO 954.

XX

KW Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;

KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;

KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;

KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;

KW anti-inflammatory; immunosuppressive; neuroprotective; nootropic;

KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;

KW antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;

KW secretory polynucleotide; secretory protein.

XX

OS Homo sapiens.

XX

PN WO200283876-A2.

XX

PD 24-OCT-2002.

XX

PF 27-MAR-2002; 2002WO-US009921.

XX

PR 29-MAR-2001; 2001US-0280067P.

PR 29-MAR-2001; 2001US-0280068P.

PR 16-MAY-2001; 2001US-0291280P.

PR 17-MAY-2001; 2001US-0291829P.

PR 17-MAY-2001; 2001US-0291849P.

PR 19-JUN-2001; 2001US-0299428P.

PR 20-JUN-2001; 2001US-0299776P.

Db 121 LPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKNVNPGKRG 180

Qy 972 PVGDCTGTLPADWKHGGYLNELGAFGLDRSYMKGSLKDLGKNSEYNSSNCSLSSSENPYA 1031
 |||

Db 181 PVGDCTGTLPADWKHGGYLNELGAFGLDRSYMKGSLKDLGKNSEYNSSNCSLSSSENPYA 240

Qy 1032 TIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPTVSVVQGVFSNN 1091
 |||

Db 241 TIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPTVSVVQGVFSNN 300

Qy 1092 GRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE 1140
 |||

Db 301 GRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE 349

RESULT 13

ABJ37901

ID ABJ37901 standard; protein; 1450 AA.

XX

AC ABJ37901;

XX

DT 22-MAY-2003 (first entry)

XX

DE NOVX protein sequence SEQ ID No 48.

XX

KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
 KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
 KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
 KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
 KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
 KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
 KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
 KW immunogen; non-human transgenic animal; gene therapy.

XX

OS Unidentified.

XX

PN WO200281517-A2.

XX

PD 17-OCT-2002.

XX

PF 22-JAN-2002; 2002WO-US002064.

XX

PR 19-JAN-2001; 2001US-0262892P.

PR 23-JAN-2001; 2001US-0263598P.

PR 24-JAN-2001; 2001US-0263799P.

PR 25-JAN-2001; 2001US-0264117P.

PR 25-JAN-2001; 2001US-0264139P.

PR 26-JAN-2001; 2001US-0264478P.

PR 30-JAN-2001; 2001US-0263351P.

PR 02-MAR-2001; 2001US-0272870P.

PR 14-MAR-2001; 2001US-0275927P.

PR 14-MAR-2001; 2001US-0275990P.

PR 15-MAR-2001; 2001US-0276449P.

PR 20-MAR-2001; 2001US-0277358P.

PR 23-MAR-2001; 2001US-0278151P.
PR 29-MAR-2001; 2001US-0279857P.
PR 20-APR-2001; 2001US-0285140P.
PR 20-APR-2001; 2001US-0285141P.
PR 30-APR-2001; 2001US-0287484P.
PR 17-MAY-2001; 2001US-0291701P.
PR 08-JUN-2001; 2001US-0296960P.
PR 10-JUL-2001; 2001US-0304353P.
PR 10-JUL-2001; 2001US-0304355P.
PR 12-JUL-2001; 2001US-0304886P.
PR 09-AUG-2001; 2001US-0311289P.
PR 13-AUG-2001; 2001US-0311975P.
PR 16-AUG-2001; 2001US-0312937P.
PR 18-OCT-2001; 2001US-0330227P.
PR 29-NOV-2001; 2001US-0334198P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;

XX

DR WPI; 2003-058504/05.

DR N-PSDB; ABT33366.

XX

PT New polypeptides, designated as NOVX, useful for diagnosing and treating
PT infections, neurological diseases, cancer, allergy, and bone,
PT immunological, skin, renal, brain, muscle and autoimmune disorders.

XX

PS Claim 1; Page 129-130; 672pp; English.

XX

CC The invention relates to a novel isolated polypeptide, designated NOVX
CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
CC the specification, or its variant, where amino acid residue(s) in the
CC variant differ from the mature form, provided that the variant differs in
CC not more than 15 % of the amino acids from the sequence of the mature
CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
CC an antibody to the polypeptides, are useful for treating or preventing a
CC NOVX-associated disorder in humans and for treating a syndrome associated
CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
CC the encoding nucleic acids, are useful for determining the presence of or
CC predisposition to a disease associated with altered levels of NOVX
CC polypeptide and polynucleotide, by measuring the level of polypeptide
CC expression or the amount of nucleic acid from a mammal and comparing it
CC with another mammal not having or not predisposed to the disease. NOVX
CC polypeptide is also useful for identifying an agent that binds to NOVX
CC and a cell expressing NOVX is useful for identifying an agent that
CC modulates the expression or activity of NOVX. The antibodies and a
CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
CC for treating a pathological state in a mammal. The antibodies are also
CC useful for determining the presence or amount of NOVX in a sample. NOVX
CC polypeptides, polynucleotides and antibodies specific for the
CC polypeptides are useful for treating or preventing disorders or syndromes
CC including trauma, viral, bacterial, fungal, protozoal, and parasitic

infections. They can also treat disorders such as e.g., Alzheimer's disease or a stroke. The NOVX encoding nucleic acids are useful for expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful for identifying a cell or tissue type in a biological sample, to amplify DNA sequences from very small biological samples such as tissues e.g. hair or skin or body fluids in forensic biology and as primers and probes for use in identifying and/or cloning NOVX homologues in other cell types. The NOVX proteins are useful as an immunogen to generate antibodies which are useful for diagnostically monitoring protein levels and modulating NOVX activity. Cells comprising NOVX nucleic acids are useful for producing non-human transgenic animals which are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. The NOVX nucleic acids can be used in gene therapy. This sequence represents a NOVX protein of the invention

XX

SQ Sequence 1450 AA;

Query Match 27.6%; Score 1858.5; DB 6; Length 1450;
Best Local Similarity 42.0%; Pred. No. 1.1e-83;
Matches 318; Conservative 83; Mismatches 324; Indels 33; Gaps 12;

Qy	94	CCPGFYESGEMC-----VPHCADKCVHGRCIAPNT-----CQCEPGWGGTNC	135
Db	624	CDPLGY--GRFCHLACPPWAF'GPGCSEEC---QCVQPHQTQSCDKRDGSCSCKAGFRGERC	678
Qy	136	SSACDGDHWPHTSRQCCKNGALCNPITGAC--HCAAGFRGWRCEDRCEQGTYGNDCHQ	193
Db	679	QAECELGYFGPGCWQACTCPVGVACDSVSGECGKRCPAGFQGEDCGQECVPVGTFGVNCSS	738
Qy	194	RCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCP-CQNGGVCHHVTGE	252
Db	739	SCSC--GGAPCHGVTGQRCRPPGRTGEDCEADCPEGRWGLGCQEICPACQHAARCDPETGA	797
Qy	253	CSCPSGWMGTVCGPCEPGRFGKNCSQECQCHNGGTDAATGQCHCSPGYTGERCQDECP	312
Db	798	CLCLPGFVGSRCQDVCAGWYGPSCQTRCSCANDGHCHPATGHCSCAPGWTGFSCQRACD	857
Qy	313	VGTYGVLCAETCQCVNG-GKCYHVSAGACLCEAGFAGERCEARLCPGEGLYGIKCDKRCPC	371
Db	858	TGHWPDCSHPCNCSAGHGSCDAISGLCLCEAGYVGPCEQQ-CPQGHFGPGCEQLCQC-	915
Qy	372	LENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCA	431
Db	916	-QHGAACDHVSGACTCPAGWRGTFCHEACPAGEFFGLDCRSACNCTAGAACDAVNGSCLCP	974
Qy	432	PGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGWHGVDCSIRCPSGTW	491
Db	975	AGRRGPRCAETCPAGLYGDNCRHSCLCQNGGTCDPVSGHCACPEGWAGLACEKECPDRV	1034
Qy	492	GFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAERCDCSHADGCH	551
Db	1035	RAGCRHSGGCLNGGLCDPHTGRCLCPAGWAGDKCQSPCLRGWPGEACAQHCSCPPGAACH	1094
Qy	552	PTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTTT	611

Db 1095 HVTGACRCPPGFTGSGCEQGCPPGRYGPGEQLCGCLNGGSCDAATGACRCPTGFLGTDC 1154
 Qy 612 QRICSPGFYGHRCSTCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAG 671
 | | :| | : | | : | | | | | | | | | |
 Db 1155 NLTCPQGRFGPNCTHVC-GCGQGA-CDPVTGTCLCPPGRAGVRCERGCPQNRFGVGCEH 1212
 Qy 672 ICTCTNNGTCNPIDRSCQCYPGWIGSDCSQCPPAHWGPNCIHTCNCHNGAFCSAYDGEC 731
 | :| | | | : : | | | | | | | | | :| | | :| | :
 Db 1213 TCSCRNGGLCHASNGSCSCGLGWTGRHCELACPPGRYGAACHLECSCHNNSTGEPATGTC 1272
 Qy 732 KCTPGWTGLYCTQRCPLGFYKGKCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKCPS 791
 :| ||: | | | |||: | | :| ||:| | | ||:| | | | ||: |
 Db 1273 RCGPGFYGQACEHPCPPGFHGAGCQGLCWCQHGAPCDPISGRCLCPAGFHGHFCERGCEP 1332
 Qy 792 GTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCD 829
 | :| | | | | | : | | :| | | | | | | :
 Db 1333 GSFGECHQRCDGAPCDPVTGLCLCPPGRSGATCN 1370

RESULT 14

ABU03489

ID ABU03489 standard; protein; 739 AA.

XX

AC ABU03489;

XX

DT 21-JAN-2003 (first entry)

XX

DE Angiogenesis-associated human protein sequence #34.

XX

KW Human; angiogenesis-associated transcript; angiogenesis;
 KW angiogenesis-associated disease; cancer; cytostatic.

XX

OS Homo sapiens.

XX

PN WO200279492-A2.

XX

PD 10-OCT-2002.

XX

PF 14-FEB-2002; 2002WO-US004915.

XX

PR 14-FEB-2001; 2001US-00784356.

PR 22-FEB-2001; 2001US-00791390.

PR 19-APR-2001; 2001US-0285475P.

PR 03-AUG-2001; 2001US-0310025P.

PR 13-NOV-2001; 2001US-0350666P.

PR 29-NOV-2001; 2001US-0334244P.

XX

PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX

PI Murray R, Glynne R, Watson SR, Aziz N;

XX

DR WPI; 2003-040681/03.

DR N-PSDB; ABX08773.

XX

PT Detecting angiogenesis-associated transcript in a cell for diagnosing and
 PT treating cancer by contacting a sample with a polynucleotide that
 PT exhibits changes in expression level as a function of time in tissue

PT undergoing angiogenesis.

XX

PS Example 2; Page 212; 291pp; English.

XX

CC The present invention relates to methods and compositions for detecting
CC an angiogenesis-associated transcript in a cell in a patient. The method
CC involves contacting a biological sample from the patient with a
CC polynucleotide that selectively hybridises to a sequence at least 80%
CC identical to any of the angiogenesis-associated human polynucleotide
CC sequences given in the specification. These angiogenesis-associated
CC polynucleotide sequences comprise genes that exhibit changes in
CC expression levels as a function of time in tissue undergoing
CC angiogenesis. The method and the polynucleotide sequences of the
CC invention are useful for diagnosing and treating angiogenesis and
CC angiogenesis-associated diseases e.g. cancer. The polynucleotide
CC sequences are also useful in the gene therapy of such disorders. The
CC angiogenesis-associated proteins encoded by the polynucleotide sequences
CC are useful as a vaccine for therapeutic and prophylactic immunisation.
CC ABU03456-ABU03569 represent angiogenesis-associated protein sequences

XX

SQ Sequence 739 AA;

Query Match 27.4%; Score 1847.5; DB 6; Length 739;
Best Local Similarity 40.0%; Pred. No. 2e-83;
Matches 364; Conservative 91; Mismatches 240; Indels 215; Gaps 24;

Qy 261 GTVCGQPCPEGRFGKNCSEQCQCHNGGTCDAATGQCHCSPGYTGERCQDECPVGTYGVL 320
||:| |||| | |||||:|||| | ||| |:|||||:|:|:|||| | :|
Db 1 GTICSLPCPEGFHGPNCSEQCRCHNGGLCDRFTGQCRCAPGYTGDRCREECPVGRFGQDC 60

Qy 321 AETCQCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHP 380
|||| | :| :| |||| | | :|| ||||:| ||: | | | :| |||
Db 61 AETCDCAPDARCFPANGACLCEHGF TGDRCTDRLCPDGFYGLSCQAPCTCDREHSLSCHP 120

Qy 381 MSGECACKPGWSGLYCNETCSPGFYGEACQOICSCQNGADCDSVTGKCTCAPGFKGIDCS 440
|:|||:| |||:|:|:|:| :| ||: | | :| | : :| | ||||: | |:
Db 121 MNGECSCLPGWAGLHCNESC PQDTHGPGCQEHCLCLHGGVCQATSGLCQCAPGYTGPHCA 180

Qy 441 TPCPLGTYGINCSSRCGCKNDVCS PVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLT CQ 500
: || |||:|:|:| | :| |||:| | || | :||: || |||| | | :||
Db 181 SLCPPD TYGVNCSARCSCE NAIACSPIDGECVCKEGWQRGNC SVPCPPGTWGFSCNASCQ 240

Qy 501 CLNGGACNTLDGTCTCAPGWRGEKCELPQDGT YGLNCAERCD CSHADGCHPTTGHC RCL 560
| : | : | ||| || | | :|| | :| || |||| |:||| | | |:
Db 241 CAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASRCD CDHSDGCDPVHGR CQCQ 300

Qy 561 PGWSGVHCDSVCAEGRWGPNC SLPCYCKNGASCSPDDGICECAPGFRGTTCQRICSPGFY 620
|| | | | || || || | |||| :| |:| | ||||| :|| | || |
Db 301 AGWMGARCHLSCPEGLWGVNCSNTCTCKNGGTCLPENGNCVCAPGFRGPSCQRSCQPGRY 360

Qy 621 GHRCSQTC PQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGT 680
| | | :|
Db 361 GKR-----CVP----- 366

Qy 681 CNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGL 740
| | | :| :| | | |||
Db 367 -----CKCANHSFCHPSNGTCYCLAGWTGP 391

QY 741 YCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQ 800
 | : | | | | | : | : | : | | | | | | | | | | | | | | | | | |
 Db 392 DCSQRCPLGTFGANCSQPCQCGPGEKC-----HPE----- 421

QY 801 ICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTALPAD--SY-QIG 857
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 422 -----TGACVCPFGHSGAPCR-----IG----IQEPFTVMPTTPVAYNSLG 458

QY 858 AIAGIILVLVLFLLALFIIYRHKQKGKESSMPAVTYTPAMRVVNADYTISGTLPHSNG 917
 | : | | : | : | : | | | | | | | | | | | | | | | | | | | | | |
 Db 459 AVIGIAVLGSLVVALVALFIGYRHWQKGKEHHHLAVAYSSG-RLDGSEYVMPDVPP---- 513

QY 918 GNANSHYFTNPSYHTLTQCATS PHVNNRDRMTVTKSKNNQLFVNLKN-VNPGKRGPVG-D 975
 : | | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 514 --SYSHYYSNPSYHTLSQCSPNPPPPNK-----VPGPLFASLQNPERPG--GAQGH 561

QY 976 CTGTLPADWKH-----GGYLNELGAFGLDRSYMGKSL-----KDLGKNSEYNSSNCS 1022
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 562 NHTTLPADWKHRREPPPGPLDR-GSSRLDRSYSYSYNGPGPFYNKGLISEELGASVAS 620

QY 1023 LSSSENPYATIKDPPVLIPKSSECGYVEMKSP-----ARRDSPYAEINNS 1067
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 621 L-SSENPYATIRDLPSLPGGPRESSYMEMKGPPSGSPPRQPPQFWDSQRRRQPPQQRDSG 679

QY 1068 TSANRNVYEVEPTVSVVQGVFSNNGRLSQDP-----YDLPKNSHIPCHYDLLPVRD 1118
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 680 T-----YE-QPSPL----IHDRDSVGSQPPLPPGLPPGHYDSPKNSHIPGHYDLPPVRH 728

QY 1119 SSSSP-KQED 1127
 | | : : | |
 Db 729 PPSPLRRQD 738

RESULT 15

ABJ37903

ID ABJ37903 standard; protein; 1403 AA.

XX

AC ABJ37903;

XX

DT 22-MAY-2003 (first entry)

XX

DE NOVX protein sequence SEQ ID No 52.

XX

KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;
 KW vulnerary; virucide; antibacterial; protozoacide; fungicide; nootropic;
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;
 KW anticonvulsant; antiaddictive; analgesic; dermatological; keratolytic;
 KW antiseborrheic; antirheumatic; antiarthritic; antiinflammatory; anti-HIV;
 KW cytostatic; antiasthmatic; antipsoriatic; hypotensive; osteopathic;
 KW antiulcer; anorectic; antidiabetic; antiallergic; haemostatic;
 KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;
 KW immunogen; non-human transgenic animal; gene therapy.

XX

OS Unidentified.

XX
 PN WO200281517-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 22-JAN-2002; 2002WO-US002064.
 XX
 PR 19-JAN-2001; 2001US-0262892P.
 PR 23-JAN-2001; 2001US-0263598P.
 PR 24-JAN-2001; 2001US-0263799P.
 PR 25-JAN-2001; 2001US-0264117P.
 PR 25-JAN-2001; 2001US-0264139P.
 PR 26-JAN-2001; 2001US-0264478P.
 PR 30-JAN-2001; 2001US-0263351P.
 PR 02-MAR-2001; 2001US-0272870P.
 PR 14-MAR-2001; 2001US-0275927P.
 PR 14-MAR-2001; 2001US-0275990P.
 PR 15-MAR-2001; 2001US-0276449P.
 PR 20-MAR-2001; 2001US-0277358P.
 PR 23-MAR-2001; 2001US-0278151P.
 PR 29-MAR-2001; 2001US-0279857P.
 PR 20-APR-2001; 2001US-0285140P.
 PR 20-APR-2001; 2001US-0285141P.
 PR 30-APR-2001; 2001US-0287484P.
 PR 17-MAY-2001; 2001US-0291701P.
 PR 08-JUN-2001; 2001US-0296960P.
 PR 10-JUL-2001; 2001US-0304353P.
 PR 10-JUL-2001; 2001US-0304355P.
 PR 12-JUL-2001; 2001US-0304886P.
 PR 09-AUG-2001; 2001US-0311289P.
 PR 13-AUG-2001; 2001US-0311975P.
 PR 16-AUG-2001; 2001US-0312937P.
 PR 18-OCT-2001; 2001US-0330227P.
 PR 29-NOV-2001; 2001US-0334198P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;
 PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;
 PI Kekuda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;
 PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shimkets RA, Gusev V;
 PI Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;
 PI Fernandes E, Smithson G, Malyankar U, Taillon B, Liu X;
 XX
 DR WPI; 2003-058504/05.
 DR N-PSDB; ABT33368.
 XX
 PT New polypeptides, designated as NOVX, useful for diagnosing and treating
 PT infections, neurological diseases, cancer, allergy, and bone,
 PT immunological, skin, renal, brain, muscle and autoimmune disorders.
 XX
 PS Claim 1; Page 133; 672pp; English.
 XX
 CC The invention relates to a novel isolated polypeptide, designated NOVX
 CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in
 CC the specification, or its variant, where amino acid residue(s) in the
 CC variant differ from the mature form, provided that the variant differs in

CC not more than 15 % of the amino acids from the sequence of the mature
 CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and
 CC an antibody to the polypeptides, are useful for treating or preventing a
 CC NOVX-associated disorder in humans and for treating a syndrome associated
 CC with a human disease (NOVX-associated disorder). NOVX polypeptides and
 CC the encoding nucleic acids, are useful for determining the presence of or
 CC predisposition to a disease associated with altered levels of NOVX
 CC polypeptide and polynucleotide, by measuring the level of polypeptide
 CC expression or the amount of nucleic acid from a mammal and comparing it
 CC with another mammal not having or not predisposed to the disease. NOVX
 CC polypeptide is also useful for identifying an agent that binds to NOVX
 CC and a cell expressing NOVX is useful for identifying an agent that
 CC modulates the expression or activity of NOVX. The antibodies and a
 CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful
 CC for treating a pathological state in a mammal. The antibodies are also
 CC useful for determining the presence or amount of NOVX in a sample. NOVX
 CC polypeptides, polynucleotides and antibodies specific for the
 CC polypeptides are useful for treating or preventing disorders or syndromes
 CC including trauma, viral, bacterial, fungal, protozoal, and parasitic
 CC infections. They can also treat disorders such as e.g., Alzheimer's
 CC disease or a stroke. The NOVX encoding nucleic acids are useful for
 CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in
 CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful
 CC for identifying a cell or tissue type in a biological sample, to amplify
 CC DNA sequences from very small biological samples such as tissues e.g.
 CC hair or skin or body fluids in forensic biology and as primers and probes
 CC for use in identifying and/or cloning NOVX homologues in other cell
 CC types. The NOVX proteins are useful as an immunogen to generate
 CC antibodies which are useful for diagnostically monitoring protein levels
 CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are
 CC useful for producing non-human transgenic animals which are useful for
 CC studying the function and/or activity of NOVX protein and for identifying
 CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic
 CC acids can be used in gene therapy. This sequence represents a NOVX
 CC protein of the invention

XX

SQ Sequence 1403 AA;

Query Match 26.2%; Score 1770; DB 6; Length 1403;
 Best Local Similarity 35.9%; Pred. No. 2.4e-79;
 Matches 317; Conservative 74; Mismatches 345; Indels 146; Gaps 13;

Qy 83 GEKTM YRRKSQCCPGFYESGEMCVPHCADKCVHGRC-IAPNTCQCEPGWGGTNCSSACDG 141
 || |: | | : | | | | | : | | || | |: :|
 Db 503 GEHTL-TEKFVCLDDSF---GHDCSLTCDDCRNGGTCLLGLDGCDCPEGWTGLICNESCP 559
 Qy 142 DHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCN 201
 | :| :|: | | :|| | : :||| | | | || | :| || | :| || |
 Db 560 DTFGKNCSFSCSQNGGTCDSVTGACRCPPGVSGTNCEDGCPKGYGKHCRKKNCANRG 619
 Qy 202 TCDHVTGECRCPPGYTGAFCEDLCP----- 226
 | : | | | | | | | |
 Db 620 RCHRLYGACLCDPGLYGRFCHLACPPWAFGPGCSEECQCVQPHQTQSCDKRDGSCSCKAGF 679
 Qy 227 -----PGKHGPQCEQRCPCQNGGVCHHVTGECs--CPSGWMGTVCQPCPEGRFG 274
 || || | | | | | | | :||| ||:|: | || | | | |
 Db 680 RGERCQAECEPGYFGPGCWQACTCPVGVACDSVSGECGKRCPAGFQGEDCGQECVPVGTFG 739

Qy	275	KNCSQECQ-----CHNGGTCD	291
Db	740	VNCSSSSCSCGGAPCHGVTGQCRCPPGRTGEDCEAGECEGLWGLGCQEICPACHNAARCDP	799
Qy	292	ATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCQVNGGKCYH-----	334
		: :	
Db	800	ETGACLCLPGFVGSRQD-CEAGWYGPSCQTMSCANDGHCHQDTGHCSCAPGWTGFSCQ	858
Qy	335	-----VSGACLCEAGFAGERCEARLCPEGLYGIKCDKR	367
		: : : : :	
Db	859	RACDTGHWGPDCSHPCNCSAGHGSDAISGLCLCEAGYVGPRCEQSECPQGHFGPGCEQR	918
Qy	368	CPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGK	427
		: : : : :	
Db	919	CQC--QHGAACDHVSGACTCPAGWRGTFCHEACPAFFGLDCRSACNCTAGAACDAVNGS	976
Qy	428	CTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGVDCSIRCP	487
		: :	
Db	977	CLCPAGRRGPRCAETCPAHTYGHNCSQACACFNGASCDPVHGQCHCAPGWMGPSCIQACP	1036
Qy	488	SGTWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHA	547
		: : : : : : :	
Db	1037	AGLYGDNCRHSCLCQNGGTCDPVSGHCACPEGWAGLACEVECLPRDVRAGCRHSGGCLNG	1096
Qy	548	DGCHPTTGHCRCCLPGWSGVHCDS--VCAEGRWGPNCSLPCYCKNGASCSPPDDGICECAPG	605
		: : : : :	
Db	1097	GLCDPHTGRCLCPAGWTGDKCQSPAACAKGTFGPHCEGRACACRWGGPCHLATGACLCPPG	1156
Qy	606	FRGTTTCQRICSPGFYGHRCSTCPQCQVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRF	665
		: : : : : : :	
Db	1157	WRGPHLSAACLRGWFGEACAQRC-SCPPGAA-CHHVTGACRCPPGFTGSGCEQACPPGSF	1214
Qy	666	GKNCAGICTCT-NNGTCPNIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFC	724
		: : : : :	
Db	1215	GEDCAQMCQCPGENPACHPATGTCSAAGYHGPSCQQRCPGPRYGPGEQLCGCLNGGSC	1274
Qy	725	SAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDDHISGQCTCRTGFMGRH	784
		: : : : : :	
Db	1275	DAATGACRCPTGFLGTDCNLTCPQGRFGPNCTHVCGCGQGAACDPVTGTCLCPPGRAGVR	1334
Qy	785	CEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGA	826
		: :	
Db	1335	CERGCQNRFVGVEHTCSCRNGGLCHASKRQLLLLWPGLDGA	1376

Search completed: March 26, 2004, 16:08:52
Job time : 78.3326 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:06:56 ; Search time 25.759 Seconds
(without alignments)
2284.780 Million cell updates/sec

Title: US-10-092-390-2
Perfect score: 6744
Sequence: 1 MVISLNSCLSFICLLCHWI.....SSPKQEDSGGSSSNSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	1037	15.4	2523	1	US-08-185-432-18	Sequence 18, Appl
2	1037	15.4	2523	4	US-08-899-232-3	Sequence 3, Appli
3	1035.5	15.4	2556	1	US-08-185-432-17	Sequence 17, Appl
4	1035.5	15.4	2556	4	US-08-899-232-2	Sequence 2, Appli
5	1034.5	15.3	2556	1	US-08-083-590A-20	Sequence 20, Appl
6	1034.5	15.3	2556	3	US-08-532-384-20	Sequence 20, Appl
7	1014.5	15.0	2471	1	US-08-185-432-16	Sequence 16, Appl
8	1014.5	15.0	2471	1	US-08-083-590A-19	Sequence 19, Appl
9	1014.5	15.0	2471	3	US-08-532-384-19	Sequence 19, Appl
10	1014.5	15.0	2471	4	US-08-899-232-1	Sequence 1, Appli
11	978.5	14.5	2703	1	US-08-185-432-19	Sequence 19, Appl

12	978.5	14.5	2703	4	US-08-899-232-4	Sequence 4, Appli
13	974	14.4	2321	4	US-09-230-652-2	Sequence 2, Appli
14	953.5	14.1	1964	4	US-09-467-997-1	Sequence 1, Appli
15	790	11.7	1055	3	US-09-214-278-2	Sequence 2, Appli
16	790	11.7	1055	4	US-09-855-722-2	Sequence 2, Appli
17	789	11.7	1212	3	US-09-214-278-3	Sequence 3, Appli
18	789	11.7	1212	4	US-09-855-722-3	Sequence 3, Appli
19	781	11.6	1257	3	US-08-611-729A-8	Sequence 8, Appli
20	779.5	11.6	1065	2	US-08-400-159-8	Sequence 8, Appli
21	779	11.6	299	3	US-09-188-930-332	Sequence 332, App
22	779	11.6	299	4	US-09-312-283C-192	Sequence 192, App
23	779	11.6	299	4	US-09-312-283C-332	Sequence 332, App
24	779	11.6	1193	2	US-08-400-159-10	Sequence 10, Appl
25	779	11.6	1193	3	US-08-611-729A-10	Sequence 10, Appl
26	779	11.6	1238	3	US-09-214-278-5	Sequence 5, Appli
27	779	11.6	1238	4	US-09-855-722-5	Sequence 5, Appli
28	769	11.4	2200	4	US-09-796-575-2	Sequence 2, Appli
29	759	11.3	299	3	US-09-188-930-192	Sequence 192, App
30	757	11.2	1036	4	US-09-068-740A-6	Sequence 6, Appli
31	757	11.2	1187	4	US-09-068-740A-7	Sequence 7, Appli
32	757	11.2	1208	4	US-09-199-865-1	Sequence 1, Appli
33	757	11.2	1218	2	US-08-400-159-6	Sequence 6, Appli
34	757	11.2	1218	3	US-08-611-729A-6	Sequence 6, Appli
35	757	11.2	1218	3	US-08-882-046-2	Sequence 2, Appli
36	757	11.2	1218	4	US-09-068-740A-11	Sequence 11, Appl
37	755.5	11.2	2199	4	US-08-793-273C-2	Sequence 2, Appli
38	755.5	11.2	2199	5	PCT-US95-11684-2	Sequence 2, Appli
39	755	11.2	1010	3	US-08-882-046-7	Sequence 7, Appli
40	755	11.2	1148	3	US-08-882-046-4	Sequence 4, Appli
41	747	11.1	1218	3	US-09-214-278-7	Sequence 7, Appli
42	747	11.1	1218	4	US-09-855-722-7	Sequence 7, Appli
43	745	11.0	1248	3	US-08-882-046-6	Sequence 6, Appli
44	744	11.0	1219	3	US-08-882-046-5	Sequence 5, Appli
45	716	10.6	3084	4	US-09-562-702A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-08-185-432-18

; Sequence 18, Application US/08185432

; Patent No. 5750652

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Busseau, Isabelle

; APPLICANT: Diederich, Robert J.

; APPLICANT: Xu, Tian

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

```

; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2523 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-18

```

```

Query Match          15.4%; Score 1037; DB 1; Length 2523;
Best Local Similarity 25.4%; Pred. No. 1.5e-60;
Matches 326; Conservative 84; Mismatches 304; Indels 568; Gaps 78;

```

```

Qy      83 GEKTMYRR----KSQC-----CP-GFYESGEMCVPHCADKCVHGR 117
      ||:  :  |:||                || ||  : ::|:  : ||:
Db      53 GERCQFPNPCTIKNQCMNFGTCEPVLQGNADFICHCPVGF--TDKVCLTPVDNACVNNP 110

Qy      118 C-----IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNP--IT 164
      |          :  |:| ||| | :| |          | |  | ||  | | |
Db      111 CRNGGTCELLNSVTEYKCRCPPGWTGDSCQQA-----DPCASN-PCANGGKCLPFEIQ 162

Qy      165 GACHCAAGFRGWRCE---DRCEQ-----GTYGNDCHQR-----CQ 196
      | |  || |  |:  : | |          |:|  |  |          |
Db      163 YICKCPPGFHGATCKQDINECSQNPCKNGGQCINEFGSYRCTCQNRFTGRNCDEPYVPCN 222

Qy      197 ---CQNGATC---DHVTGECRCPPGYTGAFCE-----LC 225
      | || ||  |  : :| | ||:|  ||:
Db      223 PSPCLNGGTCTCRQTDTSYDCTCLPGFSGQNCEENIDDCPSNNCRNGGTCVDGVNTYNCQC 282

Qy      226 PPGKHGPQCEQ---RC-----PCQNGGVCHHVTG--ECSCPSGWMGTVCQG----- 266
      ||  |  |  :  |          ||||| ||:  |  |  :||  |  :
Db      283 PPDWTGQYCTEDVDECQLMPNACQNGGTCHNTYGGYNCVCVNGWTGEDCSENIDDCANAA 342

Qy      267 -----PCPEGRFGKNC--SQEC---QCHNGGTCD--ATGQ--CHCSPG 301
      || || |  |  |          |:  |  ||          | :  | | |
Db      343 CHSGATCHDRVASFYCECPHGRTGLLCHLDNACISNPCNEGSNCDTNPVNGKAICTCPPG 402

Qy      302 YTGERCQ---DECPVGTYGVLCAETCQCVNGGKCYHVSGA--CLCEAGFAGERCEARLCP 356
      |||  |  ||| :|          | :||:| :  |:  | |  |:|| |||  :

```


Db 403 YTGPACNNDVDECSLGAN-----PCEHGGRCNTNLGSFQCNCPOQYAGPRCEIDV-- 452
 Qy 357 EGLYGIKCDKRC---PCHLENTHSCHPMSE--CACKPGWSGLYC----- 396
 | | : | : | | : |||
 Db 453 -----NECLSNPC--QNDSTCLDQIGEFQCTICMPGYEGLYCETNIDECASNPCLN 501
 Qy 397 -----NE---TCSPGFYGEACQ--QICS---CQNGADC-----DSVTGK- 427
 | | | | | : | : ||| | : ||:
 Db 502 GKCIDKINEFRCD CPTGFSGNLCQHDFDECTSTPCKNGAKCLDGPNSYTCQCTEGFTGRH 561
 Qy 428 -----CTCAPGFKGIDC-----STP----- 442
 | | : | | | |
 Db 562 CEQDINECIPDPCHYGTCKDGIATFTCLCRPGYTGRLCDNDINECLSKPCLNGGQCTDRE 621
 Qy 443 -----CPLGTYGINCSSR---CG---CKNDAVCSPVDG-SCTCKAGWHGVDCSIR---- 485
 | | | | : || : | | : || | | : | : |
 Db 622 NGYICTCPKGTGTVNCETKIDDCASNLC DN GKCIDKIDGYECTCEPGYTGKLCNININEC 681
 Qy 486 -----CPSGTWGFGC--- 495
 | | : |
 Db 682 DSNPCRNGGTCKDQINGFTCVCPDGYHDMCLSEVNECNSNPCIHGACHDGVNGYKCDCE 741
 Qy 496 -----NLTCQ---CLNGGACNTLDGT--CTCAPGWRGEKCEL-----PC-Q 530
 | | : | : ||| | : | ||| : | | : ||
 Db 742 AGWSGSNCDINNNECESNPCMNGGTCKDMTGAYICTCKAGFSGPNCQTNINECSSNPCLN 801
 Qy 531 DGT-----YGLNC-----AERCD-----CSHADGCHPT----TGHCRCLPGWS 564
 | | | | | : | : | : | | | |
 Db 802 HGTCIDDVAGYKCNMCLPYTGALCEAVLAPCAGSPCKNGGRCKESEDFFETFSCECPPGWQ 861
 Qy 565 GVHCD-----SVCAEGRWGPNC SL-----PCYCKNGASC 593
 | | : | | | | : | | | |
 Db 862 GQTCEIDMNECVNRPCRNGATCQNTNGSYKCNCKPGYTGRNCEMDIDDCQPNPCHNGGSC 921
 Qy 594 SPDDGI----CECAPGFRGTTCQR-----ICSPGFYGHRC 624
 | ||| | | ||| : | ||| | |
 Db 922 S--DGINMFFCNC PAGFRGPKCEEDINECASNPCKNGANCTDCVNSYTCTCQPGFSGIHC 979
 Qy 625 SQTCPQCVHSS----GPCHHITGL---CDCLPGFTGALC---NE----- 658
 | | | | | : | | ||| : | |
 Db 980 ESNTPDCTESSCFNGGTC--IDGINFTTCQCPPGFTGSYCQHDINECD SKPCLNGGTCQD 1037
 Qy 659 -----VCPSGRFGKNCAGI-----CTCTNNGTC---NPIDRSCQCYPGWIGSDCSQP 702
 | | | | : | | | | | : | | | | |
 Db 1038 SYGTYKCTCPQGYTGLNLCQNLVRWCDSSPCKNGGKCWQTNNFYR-CECKSGWTGVYCDVP 1096
 Qy 703 ---CPPA---HWGPNCIHTCNCHNGAFC--SAYDGECKCTPGWTGLYCTQRCPLGFYKDC 755
 | | : : | | | : | : | : | : : |
 Db 1097 SVSCEVAAKQQGV DIVHL--CRNSGMCVDTGNTHFCRCQAGYTGSYCEEQV-----DEC 1148
 Qy 756 ALICQCQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 : |||| | : : | | | : | : :
 Db 1149 S-PNPCQNGATCTDYLGGYSCECVAGYHGVNCSEEINECLSHPCQNGGTCIDLINTYKCS 1207
 Qy 789 CPSGTYGYGCRQICD-----CLNNSTC-DHITG-TCYCSPGWKGARCDQAG 832
 | | | | | | | | : | | | : | | :
 Db 1208 CPRGTQGVHCEINVDDCTPFYDSFTLEPKCFNNGKCIDRVGGYNCICPPGFVGERCE--- 1264

Qy 833 VIIVGNLNS-LSRTSTALPADS 853
 |::| | | | |
 Db 1265 ----GDVNECLSN-----PCDS 1277

RESULT 2

US-08-899-232-3

; Sequence 3, Application US/08899232
 ; Patent No. 6436650
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Qi, Huilin
 ; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
 ; FILE REFERENCE: 7326-046
 ; CURRENT APPLICATION NUMBER: US/08/899,232
 ; CURRENT FILING DATE: 1997-07-23
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 2523
 ; TYPE: PRT
 ; ORGANISM: Xenopus sp.
 US-08-899-232-3

Query Match 15.4%; Score 1037; DB 4; Length 2523;
 Best Local Similarity 25.4%; Pred. No. 1.5e-60;
 Matches 326; Conservative 84; Mismatches 304; Indels 568; Gaps 78;

Qy 83 GEKTMYYRR----KSQC-----CP-GFYESGEMCVPHCADKCVHGR 117
 ||: : |::| || | | : ::| : ||:
 Db 53 GERCQFPNPCTIKNQCMNFGTCEPVLQGNADFICHCPVGF--TDKVCLTPVDNACVNNP 110

Qy 118 C-----IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRQCKNGALCNP--IT 164
 | : |::| | | : | | | | | | | |
 Db 111 CRNGGTCELLNSVTEYKCRCPPGWTGDSCQQA-----DPCASN-PCANGGKCLPFIEIQ 162

Qy 165 GACHCAAGFRGWRCE---DRCEQ-----GTYGNDCHQR-----CQ 196
 | | | | | : : | | | : | | | |
 Db 163 YICKCPPPGFHGATCKQDINECSQNPCKNGGQCINEFGSYRCTCQNRFTGRNCDEPYVPCN 222

Qy 197 ---CQNGATC---DHVTGECRCPPGYTGAFCE-----LC 225
 | | | | | : : | | |::| | |:
 Db 223 PSPCLNGGTCRQTDTSYDCTCLPGFSGQNCENIDDCPSNNCRNGGTCVDGVNTYNCQC 282

Qy 226 PPGKHGPQCEQ---RC-----PCQNGGVCHHVTG--ECSCPSGWMGTVCQG----- 266
 || | | : | | | | | | : | | : | | :
 Db 283 PPDWTGQYCTEDVDECQLMPNACQNGGTCHNTYGGYNCVCVNGWTGEDCSENIDDCANAA 342

Qy 267 -----PCPEGRFGKNC--SQEC---QCHNGGTCD--ATGQ--CHCSPG 301
 | | | | | | | | : | | | : | | | |
 Db 343 CHSGATCHDRVASFYCECPHGRTGLLCHLDNACISNPCNEGSNCDTNPVNGKAICTCPPG 402

Qy 302 YTGERCQ---DECPVGTYGVLCAETCQCVNGGKCYHVSGA--CLCEAGFAGERCEARLCP 356
 ||| | | | : | : | | : : | | | | :
 Db 403 YTGPAACNDVDECSLGN-----PCEHGGRCNTNLGSFQCNCPPQYAGPRCEIDV-- 452

Qy 357 EGLYGIKCDKRC---PCHLENTHSCHPMSGE--CACKPGWSGLYC----- 396
 | | | : | : | | | | : | | |
 Db 453 -----NECLSNPC--QNDSTCLDQIGEFQCI CMPGYEGLYCETNIDECASNPC LHN 501

Qy 397 -----NE---TCSPGFYGEACQ---QICS---CQNGADC-----DSVTGK- 427
 | | | | | | | : | : | | | : | | :
 Db 502 GKCIDKINEFRCD CPTGFSGNLCQHDFDECTSTPCKNGAKCLDGPNSYTCQCTEGFTGRH 561

Qy 428 -----CTCAPGFKGIDC-----STP----- 442
 | | | : | | | | |
 Db 562 CEQDINECIPDPCHYGTCKDGIATFTCLCRPGYTGRLCDNDINECLSKPCLNGGQCTDRE 621

Qy 443 -----CPLGTYGINCSSR---CG---CKNDAVCSPVDG-SCTCKAGWHGVDCSIR---- 485
 | | | | : | | : | | | : | | : | | : | | : |
 Db 622 NGYICTCPKGT TGVCETKIDDCASNLC DNGKCIDKIDGYECTCEPGYTGKLCNININEC 681

Qy 486 -----CPSGTWGFGC--- 495
 | | | : |
 Db 682 DSNPCRNGGTCKDQINGFTCVCPDGYHDMCLSEVNECNSNPCIHGACHDGVNGYKCDCE 741

Qy 496 -----NLTCQ---CLNGGACNTLDGT--CTCAPGWRGEKCEL-----PC-Q 530
 | | : | : | | | : | | | | : | | : | |
 Db 742 AGWSGSNCDINNNECESNPCMNGGTCKDMTGAYICTCKAGFSGPNCQTNINECSSNPCLN 801

Qy 531 DGT-----YGLNC-----AERCD-----CSHADGCHPT----TGHCRCLPGWS 564
 | | | | | | | : | | : | : | | | | |
 Db 802 HGTCIDDVAGYKCNMCLPYTGAICEAVLAPCAGSPCKNGGRCKESEDFFETFSCECPPGWQ 861

Qy 565 GVHCD-----SVCAEGRWGPNC SL-----PCYCKNGASC 593
 | | : | | | | | : | | | | |
 Db 862 GQTCEIDMNECVNRPCRNGATCQNTNGSYKCNCKPGYTGRNCEMDIDDCQPNPCHNGGSC 921

Qy 594 SPDDGI----CECAPGFRGTTCQR-----ICSPGFYGHRC 624
 | | | | | | | : | | | | | |
 Db 922 S--DGINMFFCNC PAGFRGPKCEEDINECASNPCKNGANCTDCVNSYTCTCQPGFSGIHC 979

Qy 625 SQTCPQCVHSS----GPCHHITGL----CDCLPGFTGALC---NE----- 658
 | | | | | | | : | | | | | : | | |
 Db 980 ESNTPDCTESSCFNGGTC--IDGINTFTCQCPPGFTGSYCQHDINECD SKPCLNGGTCQD 1037

Qy 659 -----VCPSGRFGKNCAGI-----CTCTNNGTC---NPIDRSCQCYPGWIGSDCSQP 702
 | | | | | : | | | | | | : | | | | |
 Db 1038 SYGTYKCTCPQGYTGLNCQNLVRWCDSSPCKNGGKCWQTNNFYR-CECKSGWTGVYCDVP 1096

Qy 703 ---CPPA--HWGPNCIHTCNCHNGAFC--SAYDGECKCTPGWTGLYCTQRCPLGFYKDC 755
 | | | : : | | | : | : | : | | : : |
 Db 1097 SVSCEVAAKQQGV DIVHL--CRNSGMCVDTGNTHFCRCQAGYTGSYCEEQV-----DEC 1148

Qy 756 ALICQCQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 : | | | | | : : | | | : | : | : :
 Db 1149 S-PNPCQNGATCTDYLGYSCECVAGYHGVNCSEEINECLSHPCQNGGTCIDLINTYKCS 1207

Qy 789 CPSGTYGYGCRQICD-----CLNNSTC-DHITG-TCYCSPGWKGARCDQAG 832
 | | | | | | | | | | : | | | | : | | :
 Db 1208 CPRGTQGVHCEINVDDCTPFYDSFTLEPKCFNNGKCIDRVGGYNCICPPGFVGERCE--- 1264

Qy 833 VIIVGNLNS-LSRTSTALPADS 853

```

                |::|  ||      |  ||
Db      1265  ----GDVNECLSN-----PCDS  1277

```

RESULT 3

US-08-185-432-17

```

; Sequence 17, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-17

```

Query Match 15.4%; Score 1035.5; DB 1; Length 2556;
Best Local Similarity 25.9%; Pred. No. 1.9e-60;
Matches 317; Conservative 84; Mismatches 304; Indels 519; Gaps 74;

```

QY      94  CCPGFYESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGGTNCSSACDGDH 143
      |  ||  ||  ::      :  ::  |      :      ||:|  ||  |  :|  |
Db      89  CALGF--SGPLCLTPLDNACLTNPCRNGGTCDLLLTLEYKCRCPPGWSGKSCQQA----- 141

```

Qy 144 WGPHTSRCQCKNGALCNPITGA--CHCAAGFRG---WRCEDRCEQG----TYGNDCHQ- 193
 | | | | | : | | | : : | | : | |
 Db 142 --DPCASN-PCANGGQCLPFEASYICHCPPSFHGPTCWQDVNECGQKPRLCRHGGTCHNE 198

 Qy 194 -----RC-----QCQNGATC---DHVTGECRCPPGYTGAFCE--- 222
 | | | | | | | | | | : | | | |
 Db 199 VGSYRCVCRATHTGPNCEWPYVPCSPSPCQNGGTCRPTGDVTHEACLPFGFTGQNCENI 258

 Qy 223 DLCPFG--KHGPQC-----EQRCP-----CQNGGVCHHVTG- 251
 | | | | : | | | | | | | | | : | |
 Db 259 DDCPGNNCKNGGACVDGVNTYNCPCPPEWTGQYCTEDVDECQLMPNACQNGGTCHNTHGG 318

 Qy 252 -ECSCPSGWMGTVCQG-----PCPEGRFGKNC--SQEC-- 281
 | | : | | | : | | | | | : | |
 Db 319 YNCVCVNGWTGEDCSENIDDCASAACFHGATCHDRVASFYCECPHGRTGLLCHLNDACIS 378

 Qy 282 -QCHNGGTCDA--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCACETCQCVNNGKCY 333
 | : | | | : | | | | | | | : | | | |
 Db 379 NPCNEGSNCDTNPVNGKAICTCPSGYTGPAQSQDVDECSLGAN-----PCEHAGKCI 430

 Qy 334 HVSGA--CLCEAGFAGERCEARLCPEGGLYGIKCDKRC---PCHLENTHSCHPMSCGE--CA 386
 : | : | | : | | | : | | : | : | | |
 Db 431 NTLGSFECQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477

 Qy 387 CKPGWSGLYC-----NE---TCSPGFYGEACQ---QICS---C 415
 | | : | : | | | | | | | | | : | |
 Db 478 CMPGYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCQYDVDECASTPC 537

 Qy 416 QNGADC-----DSV-TGKCTCAPGFKG 436
 : | | | | | | | | | | : | |
 Db 538 KNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDDPCHYGSKDGVATFTCLCRPGYTG 597

 Qy 437 IDCST-----PCPL-----GTYGINCS-----SRCGCKNDAVCS 465
 | | | | | | | | | : | :
 Db 598 HHCETNINECSSQPCRLWGTCQDPDNAYLCFCLKGTGPNCEINLDDCASSPCDSGTCLD 657

 Qy 466 PVDG-SCTCKAGWHGVDCSIR-----CPSGTWGFGCNL-----TC---- 499
 : | | | | : | : | | : | | | | |
 Db 658 KIDGYECACEPGYTGSMCNSNIDECAGNPCHNGGTCEGDINGFTCRCPEGYHDPCLSEV 717

 Qy 500 -----QCLNGGACNTLDG-TCTCAPGWRGEKCEL-----PCQDGTYGLNC 538
 | : | : : | | | | | : : | | | : |
 Db 718 NECNSNPCVHGACWDSLNGYKCDGDPGWSGTNCDINNNECESNPCVNGGTCKDMTSGIVC 777

 Qy 539 A-----ERC---DCSHADGC-HPTTGH-CRCLPGWSGVHCDV---CAEG- 575
 | | : | | : | | : | | : | |
 Db 778 TCWEGFSGPNCQTNINECASNPCLNKGTCIDDVAGYKCNCLLPYTGATCEVVLAPCAPSP 837

 Qy 576 -RWGPNC-----SLPCYC---KNGASCSPDDGICECAPGFRGTTCQRI-----CS 616
 | | | | | : | | | : : | : | |
 Db 838 CRNGGECRQSEDYESFSCVCPTAGAKGQTCEVDINECVLSPCWHGASCQNTHGXYRCHCQ 897

 Qy 617 PGFYGHRCSQTCPCQ----VHSSGPCHH--ITGLCDCLPGFTGALCNE----- 658
 | : | | | : | | | | | | | | | |
 Db 898 AGYSGRNCETDIDDCWPNPCHNGGSCTDGINTAFCDCLPGFWGTFCCEEDINECASDPCRN 957

 Qy 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR----SCQCYPGWI 695

```

      ||:| | :|      || | | ||| :|      :| | ||:
Db      958 GANCTDCVDSYTCTCPAGFSGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGFT 1015

Qy      696 GSDC-----SQP-----CPPAHWGPNC---IHTCN---CHNGAFC 724
      || |      |:|      || : ||| |:| :| | | |
Db      1016 GSYCQHVVNECDSRPCLLGGTCDGRGLHRCTCPQGYTGPNQCQNLVHWCDSSPCKNGGKC 1075

Qy      725 ----SAYDGECKCTPGWTGLYCTQ-----R 745
      : | | |:| ||||| |
Db      1076 WQTHTQY--RCECPSGWTGLYCDVPSVSCEVAAQRQGV DVARLCQHGGLCVDAGNTHHCR 1133

Qy      746 CPLGFYKDCALI---CQ---CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
      | |:| | : | | |||| | |:| | | |:| :| ::
Db      1134 CQAGYTGSYCEDLVDECSPPCQNGATCTDYLGGYSCKCVAGYHGVCNSEEIDECLSHPC 1193

Qy      789 -----CPSGTGYGCRQICD-----CLNNSTC-DHITG-T 816
      || || | | | | | | | | | | | | : | :
Db      1194 QNGGTCLDLPTNYKCSCPWGTQGVHCEINVDDCNPPVDPVSWSPKCFNNGTCVDQVGGYS 1253

Qy      817 CYCSPGWKGARCDQAGVIVGNLN 840
      | | ||: | ||: |::|
Db      1254 CTCPPGFVGERCE-----GDVN 1270

```

RESULT 4

US-08-899-232-2

; Sequence 2, Application US/08899232

; Patent No. 6436650

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Qi, Huilin

; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON

; FILE REFERENCE: 7326-046

; CURRENT APPLICATION NUMBER: US/08/899,232

; CURRENT FILING DATE: 1997-07-23

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2556

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-899-232-2

```

Query Match          15.4%;  Score 1035.5;  DB 4;  Length 2556;
Best Local Similarity 25.9%;  Pred. No. 1.9e-60;
Matches 317;  Conservative 84;  Mismatches 304;  Indels 519;  Gaps 74;

```

```

Qy      94 CCPGFYESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGGTNCSSACDGDH 143
      | || || |:| : |:| | : |:| ||| |:| |
Db      89 CALGF--SGPLCLTPLDNACLTNPCRNGGTCDLLTLTEYKCRCPFGWSGKSCQA----- 141

Qy      144 WGPHTSRCQCKNGALCNPITGA--CHCAAGFRG---WRCEDRCEQG----TYGNDCHQ- 193
      | | | || | | : ||| | | |:| :| | :| ||
Db      142 ---DPCASN-PCANGGQCLPFEASYICHCPPSFHGPTCWQDVNECGQKPRLCRHGGTCHNE 198

Qy      194 ----RC-----QCQNGATC---DHVTGECRCPPGYTGAFCE--- 222
      || |||| || || || | |:| ||

```

Db 199 VGSYRCVCRATHTGPNCEWPHYVPCSPSPCQNGGTCRPTGDTVTHECACLPGFTGQNCENI 258
 Qy 223 DLCPPG--KHGPQC-----EQRCP-----CQNGGVCHHVTG- 251
 | || | : | | || | || | | : |
 Db 259 DDCPGNCKNGGACVDGVNTYNCPCPPEWTGQYCTEDVDECQLMPNACQNGGTCHNTHGG 318
 Qy 252 -ECSCPSGWMGTVCQG-----PCPEGRFGKNC--SQEC-- 281
 | | : | | | : | || | | | : |
 Db 319 YNCVCVNGWTGEDCSENIDDCASACFHGATCHDRVASFYCECPHGRTGLLCHLNDACIS 378
 Qy 282 -QCHNGGTCDA--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNNGKCY 333
 | : | || | : | | || | | || : | | : || |
 Db 379 NPCNEGSNCDTNPVNGKAICTCPSGYTGPAQSQDVDECSLGAN-----PCEHAGKCI 430
 Qy 334 HVSGA--CLCEAGFAGERCEARLCPEGLYGIKCDKRC---PCHLENTHSCHPMSE--CA 386
 : | : | | | : | || | : | || : | | |
 Db 431 NTLGSFECQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477
 Qy 387 CKPGWSGLYC-----NE---TCSPGFYGEACQ---QICS---C 415
 | || : | : | || | | | | | : |
 Db 478 CMPGYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLQYDVDECASTPC 537
 Qy 416 QNGADC-----DSV-TGKCTCAPGFKG 436
 : || | | | | | | : |
 Db 538 KNGAKCLDGPNTYTCVCTEGYTGTHTCEVDIDECDDPCHYGSCKDGVATFTCLCRPGYTG 597
 Qy 437 IDCST-----PCPL-----GTYGINS-----SRCGCKNDVACS 465
 | | | | | | | | : | :
 Db 598 HHCETNINECSSQPCRLWGTCQDPDNAYLCFCLKGTTPNCEINLDDCASSPCDSGTCLD 657
 Qy 466 PVDG-SCTCKAGWHGVDCSIR-----CPSGTWGFGCNL-----TC---- 499
 : || | | : | : | : | | | | |
 Db 658 KIDGYECACEPGYTGSMCNSNIDECAGNPCHNGGTCEGTINGFTCRCPEGYHDPTCLSEV 717
 Qy 500 -----QCLNGGACNTLDG-TCTCAPGWRGEKCEL-----PCQDGTYGLNC 538
 | : | : : | | | | | | : | | | : |
 Db 718 NECNSNPCVHGACWDSLNGYKCDPCDGPWSGTNCDINNECESNPCVNGGTCKDMTSGIVC 777
 Qy 539 A-----ERC---DCSHADGC-HPTTGH-CRCLPGWSGVHCDSV---CAEG- 575
 | | : | | : | | : | | |
 Db 778 TCWEGFSGPNCQTNINECASNPCLNKGTCTIDDVAGYKCNCLLPYTGTATCEVVLAPCAPSP 837
 Qy 576 -RWGPNC-----SLPCYC---KNGASCSPDDGICECAPGFRGTTCQRI-----CS 616
 | | | | | | : | | | : | : |
 Db 838 CRNGGECRQSEDIYESFSCVCPATAGAKQTCEVDINECVLSPCWHGASCQNTHGXYRCHCQ 897
 Qy 617 PGFYGHRCSQTCPCQ---VHSSGPCHH--ITGLCDCLPGFTGALCNE----- 658
 | : | | | : | | | | | | | |
 Db 898 AGYSGRNCETDIDDCWPNPCHNGGSCTDGINTAFCDCLPGFWGTFCCEEDINECASDPCRN 957
 Qy 659 -----VCPSGREGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPGWI 695
 || : | : | | | | : | : | | :
 Db 958 GANCTDCVDSYTCTCPAGFSGIHCENNTPDCTESSCFNGGTC--VDGINSF'TCLCPPGFT 1015
 Qy 696 GSDC-----SQP-----CPPAHWGPNC---IHTCN---CHNGAFC 724
 || | | : | | : || | : | | |
 Db 1016 GSYCQHVVNECDSRPCLLGGTCQDGRGLHRCCTCPQGYTGPNQNLVHWCDS SPCKNNGGKC 1075

Qy 725 ----SAYDGECKCTPGWTGLYCTQ-----R 745
: | | : | | | | | | |
Db 1076 WQTHYQY--RCECPSGWTGLYCDVPSVSCEVAAQRQGV DVARLCQHGGLCVDAGNTHHCR 1133
Qy 746 CPLGFYKDCALI---CQ---CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
| | : | | : | | | | | | : : | | : | : :
Db 1134 CQAGYTGSYCEDLVDECSPSPCQNGATCTDYLGGYSCKCVAGYHGVNCSEEIDECLSHPC 1193
Qy 789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG-T 816
| | | | | | | | | | | | | : | :
Db 1194 QNGGTCLDLPNTYKCSCPWGTQGVHCEINVDDCNPPVDPVSWSPKCFNNGTCVDQVGGYS 1253
Qy 817 CYCSPGWKGARCDQAGVIVGNLN 840
| | | : | | : | : |
Db 1254 CTCPPGFVGERCE-----GDVN 1270

RESULT 5

US-08-083-590A-20

; Sequence 20, Application US/08083590A

; Patent No. 5786158

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/083,590A

; FILING DATE: 25-JUN-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-015

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2556 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-083-590A-20

Query Match 15.3%; Score 1034.5; DB 1; Length 2556;
Best Local Similarity 25.8%; Pred. No. 2.2e-60;
Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

```
Qy      94 CCPGFYESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGGTNCSSACDGDH 143
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      89 CALGF--SGPLCLTPLDNACLTNPCRNGGTCDLLTLTEYKCRCPGWSGKSCQQA---- 141

Qy     144 WGPHTSRCQCKNGALCNPITGA--CHCAAGFRGWRCE---DRCEQG----TYGNDCHQ- 193
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     142 --DPCASN-PCANGGQCLPFEASYICHCPPSFHGPTCRQDVNECGQKPRLCRHGGTCHNE 198

Qy     194 ----RC-----QCQNGATC---DHVTGECRCPPGYTGAFCE---- 222
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     199 VGSYRCVCRATHTGPNCEPYVPCSPSPCQNGGTCTRPTGDTVTHEACALPGFTGQNCENI 258

Qy     223 DLCPPG--KHGPQC-----EQRCP-----CQNGGVCHHVTG- 251
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     259 DDCPGNNCKNGGACVDGVNTYNCPCPEWTGQYCTEDVDECQLMPNACQNGGTCHNTHGG 318

Qy     252 -ECSCPSGWMGTVCQG-----PCPEGRFGKNC--SQEC-- 281
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     319 YNCVCVNGWTGEDCSENIDDCASAACFHGATCHDRVASFYCECPHGRTGLLCHLNDAIS 378

Qy     282 -QCHNGGTCDA--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNNGKCY 333
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     379 NPCNEGSNCDTNPVNGKAICTCPSGYTGPACSQDVDECSLGAN-----PCEHAGKCI 430

Qy     334 HVSGA--CLCEAGFAGERCEARLCPEGLYGIKCDKRC---PCHLENTHSCHPMSE--CA 386
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     431 NTLGSFECQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477

Qy     387 CKPGWSGLYC-----NE---TCSPGFYGEACQ---QICS---C 415
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     478 CMPGYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLQYDVDECASTPC 537

Qy     416 QNGADC-----DSV-TGKCTCAPGFKG 436
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     538 KNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSKDGVATFTCLCRPGYTG 597

Qy     437 IDCST-----PCPL-----GTYGINCS-----SRCGCKNDAVCS 465
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     598 HHCETNINECSSQPCRLRGTCQDPDNAYLCFCLKGTTPNCEINLDDCASSPCDSGTCLD 657

Qy     466 PVDG-SCTCKAGWHGVDCSIR-----CPSGTWGFNCNL-----TC---- 499
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     658 KIDGYECACEPGYTGSMCNSNIDECAGNPCHNGGTCEGNGFTCRCPEGYHDPTCLSEV 717

Qy     500 -----QCLNGGACNTLDG-TCTCAPGWRGEKCEL----- 527
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     718 NECNSNPCVHGACRDSLNGYKCDPCPGWSGTNCDINNNECESNPCVNGGTCKDMTSGIVC 777

Qy     528 PCQDGTGYNLCAERCD-----CSHADGC-HPTTGH-CRCLPGWSGVHCDSV---CAEG- 575
```

```

      |::| | | : | : | |::| |: | |
Db      778 TCREGFSGPNCQTNINECASNPCLNKGTCIDDVAGYKCNCLLPYTGATCEVVLAPCAPSP 837

Qy      576 -RWGPNC-----SLPCYC---KNGASCSPDDGICECAPGFRGTTTCQRI-----CS 616
      | | | | | | : | | : | | |
Db      838 CRNGGECRQSEDYESFSCVCPTAGAKGQTCEVDINECVLSPCRHGASCQNTHGGRCHCQ 897

Qy      617 PGFYGHRCSTCPQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
      |: | | | | | | | | | | | |
Db      898 AGYSGRNCETDIDDC--RPNPCHNGGSCTDGINTAFCDCLPGFRGTFCEEDINECASDPC 955

Qy      659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPG 693
      |::| | : | | | | | | : | : | |
Db      956 RNGANCTDCVDSYTCTCPAGFSGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPG 1013

Qy      694 WIGSDC-----SQP-----CPPAHWGPNC---IHTCN---CHNGA 722
      : | | | | | | | | | | | | : | | |
Db      1014 FTGSYQHVVNECDSPCLLGGTCQDGRGLHRCTCPQGYTGPNQNLVHWCDSPPCKNGG 1073

Qy      723 FC----SAYDGECKCTPGWTGLYCTQ----- 744
      | : | | | | | |
Db      1074 KCWQTHQY--RCECPSGWTGLYCDVPSVSCEVAAQRQGVQDVARLCQHGGGLCVDAGNTHH 1131

Qy      745 -RCPLGFYKDCALI---CQ---CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
      || |: | | : | | | | |:: | | | |: | : | ::
Db      1132 CRCQAGYTGSYCEDLVDECSPPSCQNGATCTDYLGGYSCKCVAGYHGVNCSEEIDECLSH 1191

Qy      789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG 815
      | | | | | | | | | | | | | | | : |
Db      1192 PCQNGGTCLDLNPTYKCS CPRGTQGVHCEINVDDCNPPVDPVSRSPKCFNNGTCVDQVGG 1251

Qy      816 -TCYCSPGWKGARCDQAGVIVGNLN 840
      : | | | | : | | | : |::|
Db      1252 YSCTCPPGFVGERCE-----GDVN 1270

```

RESULT 6

US-08-532-384-20

; Sequence 20, Application US/08532384

; Patent No. 6083904

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/532,384
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/083,590
;   FILING DATE: 25-JUN-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Misrock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 7326-015
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212 790-9090
;   TELEFAX: 212 8698864/9741
;   TELEX: 66141 PENNIE
;   INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2556 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
US-08-532-384-20

```

```

Query Match          15.3%; Score 1034.5; DB 3; Length 2556;
Best Local Similarity 25.8%; Pred. No. 2.2e-60;
Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

```

```

Qy      94 CCPGFYFESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGGTNCSSACDGDH 143
      | || || :| : | : | : | : | : | : | : | : | : | : |
Db      89 CALGF--SGPLCLTPLDNACLTNPCRNGGTCDLLTLTEYKCRCPFGWSGKSCQQA----- 141

Qy     144 WGPHTSRCQCKNGALCNPIITGA--CHCAAGFRGWRCE---DRCEQG---TYGNDCHQ- 193
      | | | | | | : | | | | | : | | : | | |
Db     142 --DPCASN-PCANGGQCLPFEASYICHCPPSFHGPTCRQDVNECGQKPRLCRHGGTCHNE 198

Qy     194 ----RC-----QCQNGATC---DHVTGECRCPPGYTGAFCE--- 222
      || | | | | | | | | | | : | | |
Db     199 VGSYRCVCRATHTGPN CERPYVPCSPSPCQNGGTCRPTGDTVTHEACALPGFTGQNC EENI 258

Qy     223 DLCPPG--KHGPQC-----EQRCF-----CQNGGVCHHVTG- 251
      | || | :| | | | | | | | | | | : |
Db     259 DDCPGNCKNGGACVDGVNTYNCPCPPEWTGQYCTEDVDECQLMPNACQNGGTCHNTHGG 318

Qy     252 -ECSCPSGWMGTVCQG-----PCPEGRFGKNC--SQEC-- 281
      | | :|| | | : | | | | | | : |
Db     319 YNCVCVNGWTGEDCSENIDDCASAACFHGATCHDRVASFYCECPHGRTGLLCHLNDACIS 378

Qy     282 -QCHNGGTCD A--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY 333
      | : | | | : | | | | | | | : | | : | | |
Db     379 NPCNEGSNCDTNPVNGKAICTCPSGYTG PACSQDVDECS LGAN-----PCEHAGKCI 430

Qy     334 HVSGA--CLCEAGFAGERCEARLCPEGLYGIKCDKRC---PCHLENTHSCHPMSGE--CA 386
      : | : | | : | | | : | | : | : | | |
Db     431 NTLGSFECQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477

```

Qy 387 CKPGWSGLYC-----NE---TCSPGFYGEACQ---QICS---C 415
 | | | : | : | | | | | | | : |
 Db 478 CMPGYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLQYDVDECASTPC 537

 Qy 416 QNGADC-----DSV-TGKCTCAPGFKG 436
 : | | | | | | | : |
 Db 538 KNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCLCRPGYTG 597

 Qy 437 IDCST-----PCPL-----GTYGINCS-----SRCGCKNDAVCS 465
 | | | | | | | : | :
 Db 598 HH CETNINECSSQPCRLRGTCQDPDNAYLCFCLKGTTPNCEINLDDCASSPCDSGTCLD 657

 Qy 466 PVDG-SCTCKAGWHGVDCSIR-----CPSGTWGFGCNL-----TC----- 499
 : | | | : | : | : | | | | | |
 Db 658 KIDGYECACEPGYTGSMCNSNIDECAGNPCHNGGTCEGTINGFTCRCPEGYHDPTCLSEV 717

 Qy 500 -----QCLNGGACNTLDG-TCTCAPGWRGEKCEL----- 527
 | : | : | : | | | | | : :
 Db 718 NECNSNPCVHGACRDSLNGYKCDGDPGWSGTNCDINNNECESNPCVNGGTCKDMTSGIVC 777

 Qy 528 PCQDGTGYNCAERCD-----CSHADGC-HPTTGH-CRCLPGWSGVHCDSV---CAEG- 575
 | : | | | : | : | : | | : | | |
 Db 778 TCREGFSGPNCQTNINECASNPCLNKGTCLDDVAGYKCNCLLPYTGATCEVVLAPCAPSP 837

 Qy 576 -RWGPNC-----SLPCYC---KNGASCSPDDGICECAPGFRGTTCQRI-----CS 616
 | | | | | : | | : | : | |
 Db 838 CRNGGECRQSEDIYESFSCVCPTAGAKGQTCCEVDINECVLSPCRHGASCQNTHGGYRCHCQ 897

 Qy 617 PGFYGHRCSQTCPCQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
 | : | | | : | | | | | | | | | |
 Db 898 AGYSGRNCETDIDDC--RPNPCHNGGSCTDGINTAFCDCCLPGFRGTFCCEEDINECASDPC 955

 Qy 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPG 693
 | : | | : | | | | | : | : | | |
 Db 956 RNGANCTDCVDSYTCTCPAGFSGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPG 1013

 Qy 694 WIGSDC-----SQP-----CPPAHWGPNC---IHTCN---CHNGA 722
 : | | | : | | | : | | | : | | |
 Db 1014 FTGSYCQHVVNECDSPCLLGGTCQDGRGLHRCCTCPQGYTGPNQCQNLVHWCDS SPCKNGG 1073

 Qy 723 FC-----SAYDGECKCTPGWTGLYCTQ----- 744
 | : | : | | | | |
 Db 1074 KCWQTHQTQY--RCECPSGWTGLYCDVPSVSCVAAQRQGV DVARLCQHGGCLCVDAGNTHH 1131

 Qy 745 -RCPLGFYKDCALI---CQ---CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 | | | : | | | | | : | | : | : :
 Db 1132 CRCQAGYTGSYCEDLVDECSPPSPCQNGATCTDYLGGSCKCVAGYHGVNCSEIIDECLSH 1191

 Qy 789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG 815
 | | | | | | | | | | : |
 Db 1192 PCQNGGTCLDL PNTYKCS CPRGTQGVHCEINVDCCNPPVDPVSRSPKCFNNGTCVDQVGG 1251

 Qy 816 -TCYCSPGWKGARCDQAGV IIVGNLN 840
 : | | | : | : | : :
 Db 1252 YSCTCPPGFVGERCE-----GDVN 1270

RESULT 7

US-08-185-432-16

; Sequence 16, Application US/08185432

; Patent No. 5750652

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Busseau, Isabelle

; APPLICANT: Diederich, Robert J.

; APPLICANT: Xu, Tian

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/185,432

; FILING DATE: 21-JAN-1994

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-006

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2471 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-185-432-16

Query Match 15.0%; Score 1014.5; DB 1; Length 2471;
Best Local Similarity 23.0%; Pred. No. 4.5e-59;
Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

Qy	93	QCCPGFYESGEMCVPHCADKCVHGR-CIAPNTCQ-----CEPGWGGTNCSSACDG	141
		: : : : :	
Db	91	RCASGF--TGEDCQYSTSHPCFVSRPCLNGGTCHMLSRDTYECTCQVGFTGKEC-----	142
Qy	142	DHWGPHCTSRCQCKNGALCNPITG--ACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQN	199
		: : : : :	
Db	143	-QWTDACLSH-PCANGSTCTTVANQFSCKCLTGFTGQKCE-----TDVNECDIPGHCQH	194

Qy	200	GATCDHVTG--ECRCPPGYTGAFCEDL-----CPPGKHG	231
		: : : : :	
Db	195	GGTCLNLPGSYQCQCPQGFTGQYCDSLYVPCAPSPCVNGGTCRQTGDFTFECNCLPGFEG	254
Qy	232	PQCEQ-----RCP-----CQNGGV	245
		:	
Db	255	STCERNIDDCPNHRCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVDECLLQPNACQNGGT	314
Qy	246	CHHVTG--ECSCPSGWMGTVCQG-----PCPEGRFGKNC-	277
		: : : :	
Db	315	CANRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASFSCMCPEGKAGLLCH	374
Qy	278	-SQEC---QCHNGGTCD--ATGQ---CHCSPGYTGERCQ---DECPVGTYGVLCAETCQC	326
		: :	
Db	375	LDDACISNPCHKGALCDTNPLNGQYICTCPQGYKGADCTEDVDECAM-----ANSNPC	427
Qy	327	VNGGKCYHVSAG--CLCEAGFAGERCE-----ARLCPEGLY	360
		: : :	
Db	428	EHAGKCVNTDGAHFHCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFK	487
Qy	361	GIKCDKR-----CP-----CHLE-----NTHSC	378
		: : : :	
Db	488	GVHCELEINECQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQIDIDDCSSTPCLNGAKC	547
Qy	379	--HPMSGECACKPGWSGLYCNET-----CSPGFYGEAC--QQ	411
		: : : :	
Db	548	IDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDSYTCICNPGYMGAI CSDQ	607
Qy	412	I-----CSCQNGA-----DCDS---VTG-----K	427
		: :	
Db	608	IDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCASNP C I H G I C M D G I N R Y S	667
Qy	428	CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGCK-----NDAVCS	465
		: : : : :	
Db	668	CVCSPGFTGQRCNIDI ECASNPCKRGATCINGVNGFRICPEGPHHPSCYSQVNECLSN	727
Qy	466	P-VDGSCT-----CKAGWHGVDCSI-----RCPSGT	490
		: : : :	
Db	728	PCIHGNCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGGTC DN LVNGYRCTCKKGF	787
Qy	491	WGFGCNLTQC-----CLNGGAC-----	507
		: :	
Db	788	KGYNCQVNIDECASNPCLNQGTCTFDDISGYTCHCVLPYTGKNCQTVLAPCSPNPCENAAV	847
Qy	508	-----NTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGH--CRCL	560
		: : : : :	
Db	848	CKESPNFESYTCLCAPGWQQRCTIDIDE-----CISK-PCMNHGLCHNTQGSYMCECP	900
Qy	561	PGWSGVHCDSVCAEGRWGPNC SLPCYCKNGASCSPDDGI---CECAPGFRGTTTCQR---	613
		: : : : :	
Db	901	PGFSGMDCEEDIDDC LANP-----CQNGGSCM--DGVNTFSCLC L P G F T G D K C Q T D M N	951
Qy	614	-----ICSPGFYGHRCSTCPQCVHSS---GPCHHITGL---	644
		: : :	
Db	952	ECLSEPCKNGGTCS DYVNSYTKCQAGFDGVHCENNINECTESSCFNGGTC--VDGINSF	1009
Qy	645	-CDCLPGFTGALC---NEV-----CPSGRFGKNC---AGICT-	674

Db	1010	SCLCPVGFTGSGFCLHEINECSSHPCLNEGTCVDGLGTYRCSPLGYTGKNCQTLVNLCSSR	1069
Qy	675	--CTNNGTC--NPIDRSCQCYPGWIGSDCSQP-----	702
Db	1070	SPCKNKGTCVQKKAESQCLCPSGWAGAYCDVPNVSCDIAASRRGVLVEHLCQHSGVCINA	1129
Qy	703	-----CPPAHWGPNC---IHTC---NCHNGAFCSAYDG--ECKCTPGWTGLYCTQR--	745
Db	1130	GNTHYCQCPLGYTGSYCEEQLDECASNPCQHGATCSDFIGGYRCECVPGYQGVNCEYEVD	1189
Qy	746	-----CPLGFYG-----KDCALICQCQNGADC-DHISGQ-	773
Db	1190	ECQNQPCQNGGTCIDLNVNHFKCSCPPGTRGLLCEENIDDCARGPHCLNGGQCMDRIGGYS	1249
Qy	774	-----CTCRTGFMGRHCEQKCPST	793
Db	1250	CRCLPGFAGERCEGDINECLSNPCSSSEGLDCIQLTNDYLCVCRSAFTGRHCE-----T	1303
Qy	794	YGYGCRQICDCLNNSTCDHITG-----TCYCSPGWKGARCDQAGVVIIVGNLNSLSRTSTA	848
Db	1304	FVDVCPQM-PCLNGGTCAVASNMPDGFICRCPGFSGARCQSS----CGQVKC-----	1351
Qy	849	LPADSYQIGAIAGIIILVLVVLFLALFIIYRHKQKGKE----SSMPAVTYTPAMRVVNA	904
Db	1352	-----RKGEQCVHTASGPR-CFCPSPRDCES	1376
Qy	905	DYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLCN	964
Db	1377	GCASS---PCQHGGSC--HPQRQPPYYS-CQCA-PPFSGSR-----CELYT	1415
Qy	965	VNPGKRGVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNS---EYNSSNC	1021
Db	1416	APPST--PPATCLSQYCADKARDGVCDE-----ACNSHACQWDGGDC	1455
Qy	1022	SLSSSENPYATIKDP	1036
Db	1456	SL-TMENPWANCSSP	1469

RESULT 8

US-08-083-590A-19

; Sequence 19, Application US/08083590A

; Patent No. 5786158

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-083-590A-19

```

```

Query Match          15.0%; Score 1014.5; DB 1; Length 2471;
Best Local Similarity 23.0%; Pred. No. 4.5e-59;
Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

```

```

Qy      93 QCCPGFYESGEMCVPHCADKCVHGR-CIAPNTCQ-----CEPGWGGTNCSSACDG 141
      :| || :|| | : | | | : || | : | | |
Db      91 RCASGF--TGEDCQYSTSHPCFVSRPCLNGGTCHMLSRDTYECTCQVGF TGKEC----- 142

Qy     142 DHWGPHCTSRCQCKNGALCNPITG--ACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQN 199
      | | | | || : | : : | | | | : || | | : | | |
Db     143 -QWTDACLSH-PCANGSTCTTVANQFSCKCLTGFTGQKCE-----TDVNECDIPGHCHQ 194

Qy     200 GATCDHVTG--ECRCPPPGYTGAFCEDL-----CPPGKHHG 231
      | || : : | : || | | : || : : | | | |
Db     195 GGTCLNLPGSYQCQCPQGFTGQYCDSLYVPCAPSPCVNGGTCRQTGDFTFECNCLPGFEG 254

Qy     232 PQCEQ-----RCP-----CQNGGV 245
      || : ||| |||||
Db     255 STCERNIDDCPNHRCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVDECLLQPNACQNGGT 314

Qy     246 CHHVTG--ECSCPSGWMGTVCQG-----PCPEGRFGKNC- 277
      | : | | : || | | : |||| : | |
Db     315 CANRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASFSCMCPGKAGLLCH 374

Qy     278 -SQEC---QCHNGGTCDA--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQC 326
      | || | || || | | | | ||| : | : |
Db     375 LDDACISNPCHKGALCDTNPLNGQYICTCPQGYKGADCTEDVDECAM-----ANSNPC 427

Qy     327 VNGGKCYHVSQA--CLCEAGFAGERCE-----ARLCPEGLY 360
      : ||| : || | | | : || ||| || |
Db     428 EHAGKCVNTDGAHFHCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFK 487

```


Qy 361 GIKCDKR-----CP-----CHLE-----NTHSC 378
 | : | : | | | : | |
 Db 488 GVHCELEINECQSNPCVNNGQCVDKVNRQCLCPPGFTGPVCQIDIDDCSSTPCLNGAKC 547

 Qy 379 --HPMSGECACKPGWSGLYCNET-----CSPGFYGEAC-QQ 411
 | | | | : : | | | : | | |
 Db 548 IDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDSYTICINPGYMGAIQSDQ 607

 Qy 412 I-----CSCQNGA-----DCDS---VTG-----K 427
 | | : | | | | : | |
 Db 608 IDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCASNPCHGICMDGINRYS 667

 Qy 428 CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGCK-----NDAVCS 465
 | : | | | | | | | : | | | : : :
 Db 668 CVCSPGFTGQRCNIDIDECASNPCRKGATCINGVNGFRICPEGPHHPSCYSQVNECLSN 727

 Qy 466 P-VDGSCT-----CKAGWHGVDCSI-----RCPSGT 490
 | : | : | | | | : : : | |
 Db 728 PCIHGNCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGGTCNLDVNGYRCTCKKGF 787

 Qy 491 WGFGCNLTQ-----CLNGGAC----- 507
 | : | : | | | |
 Db 788 KGYNCQVNIDECASNPCLNQGTCTFDDISGYTCHCVLPYTGKNCQTVLAPCSPNPCENAAV 847

 Qy 508 -----NTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGH--CRCL 560
 | | | | | : : : | : | : | | | |
 Db 848 CKESPNEFESYTCLCAPGWQQRCTIDIDE-----CISK-PCMNHGLCHNTQGSYMCECP 900

 Qy 561 PGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGI----CECAPGFRGTTTCQR--- 613
 | : | : | : | | : | | | | : | | | | | |
 Db 901 PGFSGMDCEEIDDDCLANP-----CQNGGSCM--DGVNTFSCLCPLPGFTGDKCQTDMM 951

 Qy 614 -----ICSPGFYGHRCSTCPQCVHSS---GPCHHITGL--- 644
 | | | | | : | | | | : | :
 Db 952 ECLSEPCKNGGTCSQDYVNSYTCCKQAGFDGVHCENNINECTESSCFNGGTC--VDGINSF 1009

 Qy 645 -CDCLPGFTGALC----NEV-----CPSGRFGKNC---AGICT- 674
 | | | | : | | | | | | | : :
 Db 1010 SCLCPVGFTGSFCLHEINECSSHPCLNEGTCVDGLGTYRCSCLPGYTGKNCQTLVNLCSR 1069

 Qy 675 --CTNNGTC--NPIDRSCQCYPGWIGSDCSQP----- 702
 | | | | : | | | | : | |
 Db 1070 SPCKNKGTCVQKKAESQCLCPSGWAGAYCDVPNVSCDIAASRRGVLEHLCQHSGVCINA 1129

 Qy 703 -----CPPAHWGPNC---IHTC---NCHNGAFCSAYDG--ECKCTPGWTGLYCTQR-- 745
 | | : | | : | | : | | : | | : | : |
 Db 1130 GNTHYCQCPLGYTGSYCEEQLDECASNPCQHGATCSDFIGGYRCECVPGYQGVNCEYEVD 1189

 Qy 746 -----CPLGFYG-----KDCALICQCQNGADC-DHISGQ- 773
 | | | | | | | | | | | |
 Db 1190 ECQNQPCQNGGTCIDLVNHFKCSCPPGTRGLLCEENIDDCARGPHCLNGGQCMDRIGGYS 1249

 Qy 774 -----CTCRTGFMGRHCEQKCPST 793
 | | : | | | | |
 Db 1250 CRCLPGFAGERCEGDINECLSNPCSSEGLDCIQLTNDYLCVCRSAFTGRHCE-----T 1303

Qy 794 YGYGCRQICDCLNNSTCDHITG-----TCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTA 848
: | | : | | | | : | | | | : | :
Db 1304 FVDVCPQM-PCLNGGTCAVASNMPDGFICRCPGFSGARCQSS----CGQVKC----- 1351

Qy 849 LPADSYQIGAIAGIILVLVLFLLALFIIYRHKQKGKE----SSMPAVTYTPAMRVVNA 904
: | | : | | : | : | :
Db 1352 -----RKGEQCVHTASGPR-CFCPSPRDCES 1376

Qy 905 DYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKN 964
| | : | | : | | : | : | :
Db 1377 GCASS---PCQHGGSC--HPQRQPPYYS-CQCA-PPFSGSR-----CELYT 1415

Qy 965 VNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNS---EYNSSNC 1021
| | | | | : | | : | :
Db 1416 APPST--PPATCLSQYCADKARDGVCD E-----ACNSHACQWDGGDC 1455

Qy 1022 SLSSSENPYATIKDP 1036
| | : | | : |
Db 1456 SL-TMENPWANCSSP 1469

RESULT 9

US-08-532-384-19

; Sequence 19, Application US/08532384

; Patent No. 6083904

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/532,384

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/083,590

; FILING DATE: 25-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-015

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

Qy 466 P-VDGSCT-----CKAGWHGVDCSI-----RCPSGT 490
 | : | : | | | | | : : : : : | |
 Db 728 PCIHGNCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGGTCDNLVNGYRCTCKKGF 787
 Qy 491 WGFGCNLTQC-----CLNGGAC----- 507
 | : | : | | | | |
 Db 788 KGYNCQVNIDECASNPCLNQGTCTFDDISGYTCHCVLPYTGKNCQTVLAPCSPNPCENAAV 847
 Qy 508 -----NTLDGTCTCAPGWRGEKCELPQDGTGYGLNCAERCDCSHADGCHPTTGH--CRCL 560
 | | | | | : : : : : : : | : | : | | | | |
 Db 848 CKESPNFESYTCLCAPGWQQRCTIDIDE-----CISK-PCMNHGLCHNTQGSYMCECP 900
 Qy 561 PGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGI----CECAPGFRGTTTCQR--- 613
 | : | : | : | : | : | | | | | : : | | | | | | |
 Db 901 PGFSGMDCEEDIDDCLANP-----CQNGGSCM--DGVNTFSCCLCLPGFTGDKCQTDNM 951
 Qy 614 -----ICSPGFYGHRCSTCPQCVHSS----GPCHHITGL--- 644
 | | | | | : | | | | : : : :
 Db 952 ECLSEPCKNGGTCSDYVNSYTKCKQAGFDGVHCENNINECTESSCFNGGTC--VDGINSF 1009
 Qy 645 -CDCLPGFTGALC----NEV-----CPSGRFGKNC---AGICT- 674
 | | | | : | | | | | | | | : : : :
 Db 1010 SCLCPVGFTGSGFCLHEINECSSHPCLNEGTCVDGLGTYRCSCLPGYTGKNCQTLVNLCSR 1069
 Qy 675 --CTNNGTC--NPIDRSCQCYPGWIGSDCSQP----- 702
 | | | | : | | | | : | | |
 Db 1070 SPCKNKGTCVQKKAESQCLCPSGWAGAYCDVNPVNSCDIAASRRGVLEHLCQHSGVCINA 1129
 Qy 703 -----CPPAHWGPNC---IHTC--NCHNGAFCSAYDG--ECKCTPGWTGLYCTQR-- 745
 | | : | | : | | : | | : | | : | : |
 Db 1130 GNTHYCQCPLGYTGSGYCEEQLDECASNPCQHGATCSDFIGGYRCECVPGYQGVNCEYEVD 1189
 Qy 746 -----CPLGFYG-----KDCALICQCQNGADC-DHISGQ- 773
 | | | | | | | | | | | | | |
 Db 1190 ECQNQPCQNGGTCIDLNVNHFKSCPPGTRGLLCEENIDDCARGPHCLNGGQCMDRIGGYS 1249
 Qy 774 -----CTCRTGFMGRHCEQKCPST 793
 | | : | | | | | |
 Db 1250 CRCLPGFAGERCEGDINECLSNPCSSSEGLDCIQLTNDYLCVCRSAFTGRHCE-----T 1303
 Qy 794 YGYGCRQICDCLNNSTCDHITG-----TCYCSPGWKGARCDQAGVIVGNLNSLSRTSTA 848
 : | | : | | | | : | | | : | | | : | :
 Db 1304 FVDVCPQM-PCLNGGTCAVASNMPDGFICRCPGFGSARCQSS----CGQVKC----- 1351
 Qy 849 LPADSYQIGAIAGIILVLVVLFLALLFIYRHKQKGKE----SSMPAVTYTPAMRVVNA 904
 : | : : : | | : | : | : :
 Db 1352 -----RKGEQCVHTASGPR-CFCPSPRDCES 1376
 Qy 905 DYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKN 964
 | | : | : | | : : | | | : | |
 Db 1377 GCASS---PCQHGGSC--HPQRQPPYYS-CQCA-PPFSGSR-----CELYT 1415
 Qy 965 VNPGRKGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNS---EYNSSNC 1021
 | | | | : | | : : : : :
 Db 1416 APPST--PPATCLSQYCADKARDGVCDE-----ACNSHACQWDGGDC 1455

Qy 1022 SLSSSENPHYATIKDP 1036
|| : |||:| |
Db 1456 SL-TMENPWANCSSP 1469

RESULT 10

US-08-899-232-1

; Sequence 1, Application US/08899232

; Patent No. 6436650

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Qi, Huilin

; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON

; FILE REFERENCE: 7326-046

; CURRENT APPLICATION NUMBER: US/08/899,232

; CURRENT FILING DATE: 1997-07-23

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 2471

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-899-232-1

Query Match 15.0%; Score 1014.5; DB 4; Length 2471;
Best Local Similarity 23.0%; Pred. No. 4.5e-59;
Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

Qy 93 QCCPGFYESGEMCVPHCADKCVHGR-CIAPNTCQ-----CEPGWGGTNCSSACDG 141
:| || :|| | : | | | : || | : | : |
Db 91 RCASGF--TGEDCQYSTSHPCFVSRPCLNGGTCHMLSRDTYPECTCQVGFTGKEC----- 142

Qy 142 DHWGPHTSRCQCKNGALCNPITG--ACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQN 199
| | | | || : | : : | | | | : || | : | | || :
Db 143 -QWTDACLSH-PCANGSTCTTVANQFSCKCLTGFTGQKCE-----TDVNECDIPGHCQH 194

Qy 200 GATCDHVTG--ECRCPPGYTGAFCEDL-----CPPGKHG 231
| || : : | : | : || | : || : | : | | |
Db 195 GGTCLNLPGSYQCQCPQGFTGQYCDLSLYVPCAPSPCVNGGTCRQTGDETFECNCLPGFEG 254

Qy 232 PQCEQ-----RCP-----CQNGGV 245
|| : ||| ||||
Db 255 STCERNIDDCPNHRCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVDECLLQPNACQNGGT 314

Qy 246 CHHVTG--ECSCPSGWMGTVCQG-----PCPEGRFGKNC- 277
| : | | | : || | | : |||| : | |
Db 315 CANRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASFSCMCPEGKAGLLCH 374

Qy 278 -SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQC 326
| | | | | | | | | | | | : | : |
Db 375 LDDACISNPCHKGALCDTNPLNGQYICTCPQGYKGADCTEDVDECAM-----ANSNPC 427

Qy 327 VNGGKCYHVSQA--CLCEAGFAGERCE-----ARLCPEGLY 360
: ||| : | | | : || || | || |
Db 428 EHAGKCVNTDGAFHCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFK 487

Qy 361 GIKCDKR-----CP-----CHLE-----NTHSC 378

Db	488	GVHCELEINECQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQIDIDDCSSTPCLNGAKC	547
Qy	379	--HPMSGECACKPGWSGLYCNET-----CSPGFYGEAC-QQ	411
Db	548	IDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDSYTCICNPGYMGAIQSDQ	607
Qy	412	I-----CSCQNGA-----DCDS---VTG-----K	427
Db	608	IDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCASNPCIHGICMDGINRYS	667
Qy	428	CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGCK-----NDAVCS	465
Db	668	CVCSPGFTGQRCNIDIDECASNPCRKGATCINGVNGFRICPEGPHHPSCYSQVNECLSN	727
Qy	466	P-VDGSCT-----CKAGWHGVDCSI-----RCPSGT	490
Db	728	PCIHGNCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGGTCNLDLVNGYRCTCKKGF	787
Qy	491	WGFGCNLTQC-----CLNGGAC-----	507
Db	788	KGYNCQVNIDECASNPCLNQGTCTFDDISGYTCHCVLPYTGKNCQTVLAPCSNPNCEAAV	847
Qy	508	-----NTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGH--CRCL	560
Db	848	CKESPNEFESYTCLCAPGWQQRCTIDIDE-----CISK-PCMNHGLCHNTQGSYMCECP	900
Qy	561	PGWSGVHCDVCAEGRWGPNCNCLPCYCKNGASCSPDDGI---CECAPGFRGTTTCQR---	613
Db	901	PGFSGMDCEEDIDDCLANP-----CQNGGSCM--DGVNTFSCCLCLPGFTGDKCQTDNM	951
Qy	614	-----ICSPGFYGHRCSTCPQCVHSS---GPCHHITGL---	644
Db	952	ECLSEPCKNGGTCSDYVNSYTCCKQAGFDGVHCENNINECTESSCFNGGTC--VDGINSF	1009
Qy	645	-CDCLPGFTGALC---NEV-----CPSGRFGKNC---AGICT-	674
Db	1010	SCLCPVGFTGSFCLHEINECSSHPCLNEGTCVDGLGTYRCSCLPGYTGKNCQTLVNLCSR	1069
Qy	675	--CTNNGTC--NPIDRSCQCYPGWIGSDCSQP-----	702
Db	1070	SPCKNKGTCVQKKAESQCLCPSGWAGAYCDVPNVSCDIAASRRGVLEHLCQHSGVCINA	1129
Qy	703	-----CPPAHWGPNC---IHTC---NCHNGAFCSAYDG--ECKCTPGWTGLYCTQR--	745
Db	1130	GNTHYCQCPLGYTGSYCEEQLDECASNPCQHGAATCSDFIGGYRCECVPGYQGVNCEYEVD	1189
Qy	746	-----CPLGFYG-----KDCALICQCCQNGADC-DHISGQ-	773
Db	1190	ECQNQPCQNGGTCIDLNVNHFCKSCPPGTRGLLCEENIDDCARGPHCLNGGQCMDRIGGYS	1249
Qy	774	-----CTCRTGFMGRHCEQKCPST	793
Db	1250	CRCLPGFAGERCEGDINECLSNPCSSEGLDCIQLTNDYLCVCRSAETGRHCE-----T	1303
Qy	794	YGYGCRQICDCLNNSTCDHITG-----TCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTA	848

Db 1304 FVDVCPQM-PCLNGGTCAVASNMPDGFICRCPPGFSGARCQSS-----CGQVKC----- 1351
 Qy 849 LPADSYQIGAIAGIILVLVLLFLLALFIIYRHKQKGKE----SSMPAVTYTPAMRVVNA 904
 :||: :| | :|: | :
 Db 1352 -----RKGEQCVHTASGPR-CFCPSPRDCES 1376
 Qy 905 DYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKN 964
 | | :||: | | :| || | :| |
 Db 1377 GCASS---PCQHGGSC--HPQRQPPYYSCQCA-PPFSGSR-----CELYT 1415
 Qy 965 VNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNS---EYNSSNC 1021
 | | | | | | | | | | | | :| :|
 Db 1416 APPST--PPATCLSQYCADKARDGVCD-----ACNSHACQWDGGDC 1455
 Qy 1022 SLSSSENPHYATIKDP 1036
 || :||:| |
 Db 1456 SL-TMENPWANCSSP 1469

RESULT 11

US-08-185-432-19

; Sequence 19, Application US/08185432

; Patent No. 5750652

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Busseau, Isabelle

; APPLICANT: Diederich, Robert J.

; APPLICANT: Xu, Tian

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/185,432

; FILING DATE: 21-JAN-1994

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-006

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 19:

```

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2703 amino acids
;   TYPE: amino acid
;   TOPOLOGY: unknown
;   MOLECULE TYPE: protein
US-08-185-432-19

```

Query Match 14.5%; Score 978.5; DB 1; Length 2703;
Best Local Similarity 26.8%; Pred. No. 1.2e-56;
Matches 290; Conservative 103; Mismatches 296; Indels 395; Gaps 70;

Qy	7	SCL----	SFICLLLLCHWIGTASPLNLED--	PNVCSHWESYSVTVQESYPHPFDQIYYTSC	60		
Db	502	SCLDDPGTFR	CVCMPGFTGTQCEIDIDECQSNPC-----	LNDGTC	541		
Qy	61	TDILNWFKCT	RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMC---	VPHCADKCVHGR	117		
Db	542	HDKINGFKCS	-----	CALGF--TGARCQINIDDCQSQPCRNR	576		
Qy	118	-----	CIAPNTCQCEPGWGGTNCS---	SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH	168		
Db	577	GICHDSIAGYS	CECPPGYTGTSC	ININDCSN-----PCHRGKCIDDVNSFKCL	626		
Qy	169	CAAGFRGW-----	RCEDR-----	CEQGTYG-----NDCHQRCQ	196		
Db	627	CDPGYTG	YICQKQINECESNPCQFDGHCCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P		685		
Qy	197	CQNGATC-DHVTG-	ECRCPPGYTGAFCE	DLCPGKHGPQCEQRCPCQNGGVC-HHVTG-E	252		
Db	686	CNNGATCIDG	INSYKQCQVPGFTGQHCE-----	KNVDECIS-SPCANNGVCIDQVNGYK	738		
Qy	253	CSCPSGWMGTVC	-----GQP-----	CPEGRFGKNCS---QECQ--	282		
Db	739	CECPRGFYDA	HCLSDVDE	CASNPCVNEGRCE	GINEFICHCPPGYTGKRCELDIDECSSN	798	
Qy	283	-CHNGGTC-DAATG-	QCHCSPGYTGERCQ---	DECPVGTYGVLCAETCQCVNGGKCY-HV	335		
Db	799	PCQHGGTCYD	KLNAFSCQCM	PGYTQKCE	TNIDDC-----VTNPCGNGGT	848	
Qy	336	SG-ACLCEAGF	AGERCEARLCPEG	LYGIKCDK-RCPC	HLENTHSCHPMSG---ECACKP	389	
Db	849	NGYKCVCKVP	PFTGRDCESKMDP-----	CARNRC---KNEAKCTPSSN	FLDF	897	
Qy	390	GWSGLYCNETC	SPGFYGEACQQICSCQNGAD	CDSVTG--KCTCAPGFKGIDC-----	439		
Db	898	GYTGRYCD	EDI-----DECSLSSPC	RNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCA	951		
Qy	440	STPCP-----	LGTYGINC----	SSRC-----GCKND	AVCSPVDGS--CTCK	474	
Db	952	SFPCQNGRTCL	DGIGDYSLCVDG	FDGKHCE	TDINECLSQCQNGATCSQYVNSYTCTCP	1011	
Qy	475	AGWHGVDCSIR	-----	CPSGTWGF	GCN-----LTCQ-----CLN	503	
Db	1012	LGFSGINCQ	TNDE	CTESSCLN	GGSCIDGINGYNCSCLAGYSGANCQYKLNKCD	SNPCLN	1071
Qy	504	GGACNTLDG--	TCTCAPGWRGEKC-----	ELPCQDGT	YGLNCAERCD	CSHADGCHPT	553

Db 1072 GATCHEQNNEYTCHCPSGFTGKQCSEYVDWCGQSPCENG-----ATCSQMK--HQF 1120

Qy 554 TGHCRCLPGWSGVHCD---SVCAEGRWGPNCSLPCYCKNGASCSP--DDGICECAPGFRG 608
: |:| |:| || | : || | | : | : |:| |

Db 1121 S--CKCSAGWTGKLCQVQISQDAADRKGLSLRQLCNNG-TCKDYGNSHVVCYSQGYAG 1177

Qy 609 TTCQR-----ICSPGFYGHRCSTCPQCV---HSSGPCH 639
: ||: | || | | | : | ||

Db 1178 SYCQKEIDECQSQPCQNGGTCRDLIGAYECQCRQGFQGNCELNIDDCAPNPCQNGGTCH 1237

Qy 640 H--ITGLCDCLPGFTGALC---NEVCPSGRFGKNCAGICTCTNNGTCNPIDR---SCQC 690
: | | | | | : | | | | | | | | | | |

Db 1238 DRVMNFSCSCPPGTMGIIICEINKDDCKPG-----ACHNNGSC--IDRVGGFECVC 1285

Qy 691 YPGWIGSDC-----SQPCP-----PAHWGPNCIHTCN----- 717
| | : | | | | | | | | | | | |

Db 1286 QPGFVGARCEGDINECLSNPCSNAGTLDCVQLVNNYHCNCRPGHMGRHCEHKVDFCAQSP 1345

Qy 718 CHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGADCDHIS---GQC 774
| | | | : : | | | | | | | | | | |

Db 1346 CQNGGNCNIRQ-----SGHHCI--CNNGFYGKNCEL-----SGQDCDSNPCRVGNC 1389

Qy 775 TCRTGFMGRHCEQKCPSTYGYGCR---QICD---CLNNSTCDHITG--TCYCSPGWKG 825
| | | | | | | | | | : | : | | | |

Db 1390 VVADEGFGYRCE--CPRGTLGEHCEIDTLDECSPNPCAQGAACEDLLGDYECLCPKWK 1447

Qy 826 ARCD 829
| | |

Db 1448 KRCD 1451

RESULT 12

US-08-899-232-4

; Sequence 4, Application US/08899232

; Patent No. 6436650

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Qi, Huilin

; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON

; FILE REFERENCE: 7326-046

; CURRENT APPLICATION NUMBER: US/08/899,232

; CURRENT FILING DATE: 1997-07-23

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 2703

; TYPE: PRT

; ORGANISM: Drosophila sp.

US-08-899-232-4

Query Match 14.5%; Score 978.5; DB 4; Length 2703;
Best Local Similarity 26.8%; Pred. No. 1.2e-56;
Matches 290; Conservative 103; Mismatches 296; Indels 395; Gaps 70;

Qy 7 SCL----SFICLLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
| | | : | | : : | | : : : | | : : |

Db 502 SCLDDPGTFRVCMPGFTGTQCEIDIDECQSNPC-----LNDGTC 541

Qy 61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMC---VPHCADKCVHGR 117
 | : | | | : | | : | | : | : |
 Db 542 HDKINGFKCS-----CALGF--TGARCQINIDDCQSQPCNR 576

 Qy 118 -----CIAPNTCQCEPGWGGTNCS---SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH 168
 | | : | : | | | : | | : | | : | : |
 Db 577 GICHDSIAGYSCECPPGYTGTSC EININDCDSN-----PCHRGKCIDDVNSFKCL 626

 Qy 169 CAAGFRGW-----RCEDR-----CEQGTYG-----NDCHQRCQ 196
 | | : | : | | | : | | | | : | | |
 Db 627 CDPGYTGYICQKQINECESNPCQFDGHCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P 685

 Qy 197 CQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVC-HHVTG-E 252
 | | | | | | : : | : | | | | | | : : | | | | | | :
 Db 686 CNNGATCIDGINSYKQCVPFGFTGQHCE-----KNVDECIS-SPCANNGVCIDQVNGYK 738

 Qy 253 CSCPSGWMGTVC-----GQP-----CPEGRFGKNCS---QECQ-- 282
 | | | | : | | | | | | | | | | | |
 Db 739 CECPRGFYDAHCLSDVDECASNPCVNEGRCEDEGINEFICHCPPGYTGKRCELDIDECSN 798

 Qy 283 -CHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HV 335
 | : | | | | | | | | | | : | : | | | | | | |
 Db 799 PCQHGGTCYDKLNAFSCQCMPPGYTGQKCE TNIDDC-----VTNPCGNGGT CIDKV 848

 Qy 336 SG-ACLCEAGFAGERCEARLCPEGLYGIKCDK-RCPCHELENTHSCHPMSG---ECACKP 389
 : | | : | : | | | : : | | : | | | | | | | |
 Db 849 NGYKCVCKVPFTGRDCESKMDP-----CARNRC---KNEAKCTPSSNFLDF SCTCKL 897

 Qy 390 GWSGLYCNETCSPGFYGEACQQICSCQNGADCD SVTG--KCTCAPGFKGIDC----- 439
 | : | | | : | : | | | : | | | : | | | : | | | |
 Db 898 GYTGRYCDEDI-----DECSLSSPCRNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCA 951

 Qy 440 STPCP-----LGTYGINC-----SSRC-----GCKNDAVCSPVDGS--CTCK 474
 | | | : | | | | | | | : | | | | | | | | |
 Db 952 SFPCQNGRTCLDGIGDYSCLCVDGFDGKHCE TDINECLSQCQNGATCSQYVNSYTCTCP 1011

 Qy 475 AGWHGVDCSIR-----CPSGTWGF GCN-----LTCQ-----CLN 503
 | : | : | | | | : | : | | | | | | | | |
 Db 1012 LGFSGINCQTND EDC TESSCLNGGSCIDGINGYNCSCLAGYSGANCQYKLNKCD SNPCLN 1071

 Qy 504 GGACNTLDG--TCTCAPGWRGEKC-----ELPCQDGT YGLNCAERCD CSHADGCHPT 553
 | | : : | | | | : | : | : | | | | | | | |
 Db 1072 GATCHEQNEYTCHCPSGFTGKQCSEYVDWCGQSPCENG-----ATCSQMK--HQF 1120

 Qy 554 TGHCRCLPGWSGVHCD---SVCAEGRWGPNCSLPCYCKNGASCSP--DDGICECAPGFRG 608
 : | : | | : | | | : | | | | | : | : | : | : |
 Db 1121 S--CKCSAGWTGKLC DVQTISCQDAADRKGLSLRQLCNGG-TCKDYGN SHVCYCSQGYAG 1177

 Qy 609 TTCQR-----ICSPGFYGHRC SQTCPQCV----HSSGPCH 639
 : | | : | | | | | | | | : | | | | |
 Db 1178 SYCQKEIDECQSQCQNGGTCDRLIGAYECQRQGFQGNCELNIDDCAPNPCQNGGTCH 1237

 Qy 640 H--ITGLCDCLPGFTGALC---NEVCPSGRFGKNACIGICTCTNNGTCNPIDR----SCQC 690
 : | | | | | : | : | | | | | | : | | | | |
 Db 1238 DRVMNFSCSPPGTMGIIICEINKDDCKPG-----ACHNNGSC--IDRVGGFECVC 1285

QY 691 YPGWIGSDC-----SQPCP-----PAHWGPNCIHTCN----- 717
 ||::||: | | || | | : | | :
 Db 1286 QPGFVGARCEGDINECLSNPCSNAGTLDVCQLVNNYHCNCRPGHMGRHCEHKVDFCAQSP 1345

QY 718 CHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCONGADCDHIS---GQC 774
 | || | : : | : | | |||: | | : | || | |
 Db 1346 CQNGGNCNIRQ-----SGHHCII--CNNGFYGKNCEL-----SGQDCDSNPCRVGNC 1389

QY 775 TCRTGFMGRHCEQKCPSTYGYGCR---QICD---CLNNSTCDHITG--TCYCSPGWKG 825
 | || | || | | | | | | : | : | | | | ||
 Db 1390 VVADEGFGYRCE--CPRGTLGEHCEIDTLDECSPNPCAQGAACEDLLGDYECLCPKWK 1447

QY 826 ARCD 829
 |||
 Db 1448 KRCD 1451

RESULT 13

US-09-230-652-2

; Sequence 2, Application US/09230652A
 ; Patent No. 6537775
 ; GENERAL INFORMATION:
 ; APPLICANT: Tournier-Lasserve, Elisabeth
 ; APPLICANT: Joutel, Anne
 ; APPLICANT: Bousser, Marie-Germaine
 ; APPLICANT: Bach, Jean-Francois
 ; TITLE OF INVENTION: GENE INVOLVED IN CADASIL, METHOD OF DIAGNOSIS AND
 ; TITLE OF INVENTION: THERAPEUTIC APPLICATION
 ; FILE REFERENCE: 03715.0048-00000
 ; CURRENT APPLICATION NUMBER: US/09/230,652A
 ; CURRENT FILING DATE: 1999-05-17
 ; EARLIER APPLICATION NUMBER: FR 96 09733
 ; EARLIER FILING DATE: 1996-08-01
 ; EARLIER APPLICATION NUMBER: FR 97 04680
 ; EARLIER FILING DATE: 1997-04-16
 ; EARLIER APPLICATION NUMBER: PCT/FR97/01433
 ; EARLIER FILING DATE: 1997-07-31
 ; NUMBER OF SEQ ID NOS: 163
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2321
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: human ADNc No. 6537775ch 3
 US-09-230-652-2

Query Match 14.4%; Score 974; DB 4; Length 2321;
 Best Local Similarity 25.0%; Pred. No. 2.1e-56;
 Matches 304; Conservative 93; Mismatches 282; Indels 537; Gaps 70;

QY 59 SCTDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKC----- 113
 : | | : | : | | | : : | : | : |
 Db 250 TCVDGVNTYNC-----QCPPEW---TGQFCTED-VDECQLQPN 283

QY 114 -VH--GRC---IAPNTCQCEPGWG GTNCSSACDG-----DHWGPHCTSR-----CQC--- 154
 | | | : : | | | | : | | | | |

Db	284	ACHNGGTCFNTLGGHSCVCVNGWTGESCSQNIIDCATAVCFHGATCHDRVASFYCACPMG	343
Qy	155	KNGALC-----NPITG--ACHCAAGFRGWRCE---DRCE-----	183
Db	344	KTGLLCHLDDACVSNPCHEDAICDTNPVNGRAICTCPPGFTGGACDQDVDECSIGANPCE	403
Qy	184	-----QGTGYNDCHQ-----RCQ-----CQNGATCDHVTGE--CRCPPGYTG	218
Db	404	HLGRCVNTQGSFLCQCGRGYTGPRCETDVNECLSGPCRNQATCLDRIGQFTCICMAGFTG	463
Qy	219	AFCE---DLCPGKGHPQCEQRCPCQNGGVC--HHVTG--ECSCPSGWMGTVC-----	264
		:	
Db	464	TYCEVDIDEC-----QSSPCVNGGVCKDRVNGFSCTCPSGFSGSTCQLDVDECAS	513
Qy	265	-----GQP-----CPEGFRFGKNCSQ---ECQ---CHNGGTCDATGQCHCSPG	301
Db	514	TPCRNGAKCVDQPDGYECRAEGFEGTLCDRNVDDCSPDPCHHGRCVDGIASFSCACAPG	573
Qy	302	YTGERCQDE-----CPVGTGYVLC-----AETCQ---	325
Db	574	YTGTRCESQVDECRSQPCRHHGKCLDLVDKYLRCRCPSGTTGVNCEVNIDDCASNPTFGV	633
Qy	326	CVNGGKCYHVSAGACLCEAGFAGERCEARL-----CPEGLYGIKCDKRCP---	369
		:	
Db	634	CRDGINRYD---CVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFRC--LCPPGS	687
Qy	370	----CHLENTHSC-----HPMSGECACKPGWSGLYCNE-----	398
		:	
Db	688	LPPLC-LPPSHPCAHEPCSHGICYDAPGGFRCVCEPGWSGPRCSQSLARDACESQPCRAG	746
Qy	399	-----TCSPGFYGEACQOI--CS---CQNGADCDSVTGK---CTCAPGFKG--	436
Db	747	GTCSSDGMGFHCTCPPGVQGRQCELLSPCTPNPCEHGGRCESAPGQLPVCSCPPQGWQGPR	806
Qy	437	-----IDCSTPCPLGTGYGINCSSRCGKNDVAVCSPVDGSCTCKAGWHGVDCSIRCPSGTW	491
		: :	
Db	807	CQQDVDEACAGPAPCGPHGI-CTNLAG-----SFSCTCHGGYTGPSCDQDIND---	852
Qy	492	GFGCNLTQCLNGGACNTLDG----TCTCAPGWRGEKC-----ELPCQDGTYGLNCA	539
		:	
Db	853	---CDPN-PCLNGGSCQ--DGVSFSCSCLPGFAGPRCARDVDECLSNPCGPGT----CT	902
Qy	540	ERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGI	599
		: :	
Db	903	D-----HVASFTCTCPPGYGGFHCEQDL-----PDCS-PSSCFNGGTCV--DGV	943
Qy	600	----CECAPGFRGTTTCQR-----ICSPGFYGHRC----SQTCPQC-----	631
Db	944	NSFSLCRPGYTGAHCQHEADPCLSRPCLHGGVCSAAHPGFRCTCLESFTGPQCQTLVDW	1003
Qy	632	-----VHSSGPCHHITGLCDCLPGFTGALCNE-----	658
		:	
Db	1004	CSRQPCQNGGRCVQTGAYCLCPPGWSGRLCDIRSLPCREAAQIGVRLEQLCQAGGQCVD	1063
Qy	659	-----VCPSGRFGKNC---AGIC---TCTNNGTCNPI--DRSCQCYPGWIGSDC----	699
Db	1064	EDSSHYCVCPEGRTGSHCEQEVDPCLAQPCQHGGTCRGYMGGYMCECLPGYNGDNCEDDV	1123

QY 700 ----SQPC-----PPAHWGPNCIHTCNCH 719
 |||| || || |:
 Db 1124 DECASQPCQHGGSCIDLVARLYLCSCPPGTLGVLCEINEDDCGPGPPLDSGPRCLHNGTCV 1183

QY 720 N--GAFCSAYDGECKCTPGWTGLYCTQ-----RCPLG 749
 : || | | | |:
 Db 1184 DLVGGF-----RCTCPPGYTGLRCEADINECRSGACHAAHTRDCLQDPGGGFRCLCHAG 1237

QY 750 FYGKDCALI---CQ---CQNGADCDHISG-----QCTCRTGFMGRHCEQ----- 787
 | | | : | : ||: | | | | |:
 Db 1238 FSGPRCQTVLSPCESQPCQHGGQCRPSPGPGGGLTFTCHCAQPFWGPRCERVARS CRELQ 1297

QY 788 -----KCPSTYGYGCRQI-----CDCLNNSTCDHIT--- 814
 || | | | | |:
 Db 1298 CPVGVPCQQTTPRGPRCACPPGLSGPSCRSEFPGSPPGASNASCAAAPCLHGGSCRPAFLAP 1357

QY 815 -GTCYCSPGWKGARCD 829
 | : || | |:
 Db 1358 FFRCAQAQGTGPRCE 1373

RESULT 14

US-09-467-997-1

; Sequence 1, Application US/09467997

; Patent No. 6379925

; GENERAL INFORMATION:

; APPLICANT: Kitajewski, Jan

; APPLICANT: Uyttendaele, Hendrik

; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION

; FILE REFERENCE: 53863-A-PCT-US

; CURRENT APPLICATION NUMBER: US/09/467,997

; CURRENT FILING DATE: 1999-12-20

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1964

; TYPE: PRT

; ORGANISM: mouse

US-09-467-997-1

Query Match 14.1%; Score 953.5; DB 4; Length 1964;
 Best Local Similarity 25.0%; Pred. No. 4e-55;
 Matches 297; Conservative 75; Mismatches 303; Indels 513; Gaps 61;

QY 109 CADKCVHGRCI----APNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALC---- 160
 ||: | | : || | | : | | : | | | |
 Db 32 CAN---GGTCLRLSRGQGICQCAPGFLGETC-----QFPDPCRDTQLCKNGGSCQALL 81

QY 161 -----NPITG--ACHCAAGFRGWRCEDR----- 181
 : | : | | | | |:
 Db 82 PTPPSSRSPTSPLTPHFSTCPSGFTGDRQCQTHLEELCPPSFCSNGGHICYQASGRPQCS 141

QY 182 CEQGTYGNDCHQR--CQ---CQNGATC--DHVTGECRCPPGYTGAFCE----- 223
 || | | | | | | | : : || | | : | |
 Db 142 CEPGWTGEQCQLRDFCSANPCANGGVCLATYPQIQRCRPPGFEGHTCERDINECFLEPGP 201

Qy 224 -----LCPPGKHGPQCEQR---CP---CQNGGVC-----HHVTGECSC 255
 ||| |: |||: | || | ||| | | |
 Db 202 CPQGTSCHTLGSYQCLCPVGQEGPQCKLRKGACPPGSCLNGGTCQLVPEGHSTFHLCLC 261
 Qy 256 PSGWMGTVCQG-----PCPEGRFGKNCSQ---ECQ----- 282
 | |: | | |||: | :||: ||:
 Db 262 PPGFTGLDCENPDDCVRHQCNATCLDGLDITYTCPCKTWKGWDCSEIDICEARGPP 321
 Qy 283 -CHNGGTCDAATG--QCHCSPGYTGERCQDE-----CPVG 314
 | |||| | | | : | ::
 Db 322 RCRNGGTCQNTAGSFHCVCVSGWGGAGCEENLDDCAAATCAPGSTCIDRVGSFSLCPPG 381
 Qy 315 TYGVLC-----AETCQ----- 325
 |:| : ||
 Db 382 RTGLLCHLEDMCLSQPCHVNAQCSTNPLTGSTLCICQPGYSGSTCHQDLDECQMAQQGPS 441
 Qy 326 -CVNGGKCYHVSGA--CLCEAGFAGERCEAR-----LCPEG 358
 | :|| | : |: ||| |: | |||| ||| |
 Db 442 PCEHGGSCINTPGSFNCLCLPGYTGSRCADHNECLSQPCHPGSTCLDLLATFHCLCPPG 501
 Qy 359 LYGIKCD---KRC---PCHLENTHSCHPMMSG--ECACKPGWSGLYCNE----- 398
 | | |: | || | :|| : :| | ||:| | :
 Db 502 LEGRLC EVEVNECTSNPC--LNQAACHDLLNGFQCLCLPGFTGARCEKDMDECSSTPCAN 559
 Qy 399 -----TCSPGFYGEACQQICS-----CQNGADCDSVTGK--CTCAPGFKGI 437
 | ||| | :| | || | : | | ||| |
 Db 560 GGRCRDQPGAFYCECLPGFEGPHCEKEVDECLSDPCPVGASCLDLPGAFFCLCRPGFTGQ 619
 Qy 438 DCSTP-----CPLGTYGI-----NCSSRCG----- 457
 | | || : | || |
 Db 620 LCEVPLCTPNMCQPGQQCQGQEHRAPCLCPDGSPGCVPAEDNCPCHHGHQCQRLCVCDEG 679
 Qy 458 -----CKNDAVCS PVDG--SCTCKAGWHGVDCS---IRCPSGTW---- 491
 | : | | :|| ||: |: || | ||
 Db 680 WTGPECETELGGCISTPCAHHGTCHPQPSGYNCTCPAGYMGLTCSSEVTACHSGPCLNGG 739
 Qy 492 -----GFGCN-----LTCQCLNGGACNTLDGT--CTCAPGWRGEKC 525
 |: | : : |||| | || | || |:| |
 Db 740 SCSIRPEGYSCTCLPSHTGRHCQTAVDHCVSASCLNGGTCVKNKPGTFFCLCATGFQGLHC 799
 Qy 526 E-----LPCQDGTYGLNC-----AERCDCSHADGCHPT 553
 | ||| | | | | | :
 Db 800 EEKTNPSCADSPCRNKATCQDTPRGARCLCSPGYTGSSCQTLIDLCAKPCPHTARCLQS 859
 Qy 554 --TGHCRCLPGWSGVHCD--SVCAEGRWGPNC SLPCYCKNGASCSPDDG---ICECAPGF 606
 : | || ||:| || | : : || | | | |||
 Db 860 GPSFQCLCLQGTGALCDFPLSCQKAAMSQGIETISGLCQNGGLCI-DTGSSYFCRCPPGF 918
 Qy 607 RGTTCQRICSPGFYGHRCSTCPQC VHS SGPHHITGLCDCLPGFTGALCNEVCPSGRFG 666
 :| || :| | ||| : |:| :| :| | |
 Db 919 QGKLCQDNVNP-----C--EPNPCHHGS---TCVPQPSGYVCQ--CAPGYEG 958
 Qy 667 KNCAGI---C---TCTNNGTC--NPIDRSCQCYPGWIGSDC-----SQPCPPAHWGP 710
 :||: : | | :|| | | ||:| | :|| :
 Db 959 QNCSKVLDACQSQPCHNHGTCTSRPGGFHCACPPGFVGLRCEGDVDECLDRPCHPS---- 1014
 Qy 711 NCIHTCNCHNGAFCSAYDGECKCTPGWTGLYC----- 742

```

      |  |  :  ::  |:  |  |  |
Db      1015 ---GTAACH--SLANAF--YCQLPGHTGQRCEVEMDLCSQPCSNNGGSCEITTGPPPGF 1067
Qy      743 TQRCPLGFYKDC---ALIC---QCQNGADC-----DHISGQCTCRTGFMGRHC-EQKCP 790
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      1068 TCHCPKGFEGPTCSHKALSCGIHHCHNGGLCLPSPKPGSPPLCACLSGFGGPDCLTPPAP 1127
Qy      791 SGTYGYGCRQICDCLNNSTCDHITG-----TCYCSPGWKGARCDQAG 832
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      1128 P-----GCGPPSPCLHNGTCTETPGLGNPGFQCTCPPDSPGPRCQRP 1170

```

RESULT 15

US-09-214-278-2

; Sequence 2, Application US/09214278

; Patent No. 6291210

; GENERAL INFORMATION:

; APPLICANT: Sakano, Seiji

; APPLICANT: Itoh, Akira

; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE

; FILE REFERENCE: KP-8576

; CURRENT APPLICATION NUMBER: US/09/214,278

; CURRENT FILING DATE: 1999-01-26

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 1055

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-214-278-2

```

Query Match          11.7%; Score 790; DB 3; Length 1055;
Best Local Similarity 27.7%; Pred. No. 1.5e-44;
Matches 291; Conservative 85; Mismatches 377; Indels 296; Gaps 69;

```

```

Qy      15 LLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDIILNWFKCTRHRV 74
      |  |  :  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      135 LLIERVSHAGMINPEDRWKSLHFSGHVAHLELQIRVRCDENYYSATCN-----KFCRPRN 189
Qy      75 SYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPH-CADKC--VHGRCIAPNTCQCEPGWG 131
      :  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      190 DF---FGHYTCDQYGKA-CMDGW--MGKECKEAVCKQGCNLLHGGCTVPGECRCQSYGWQ 243
Qy      132 GTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 191
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      244 GRFCD-----ECVPYPGCVHGSCEVP--WQCNCETNWGGLLCDKDL-----NYC 285
Qy      192 HQRCQCQNGATCDHVTGE---CRCPPGYTGAFCEDLCPGKGHPQCEQRC---PCQNGGV 245
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      286 ESHHPCTNGGTCINAEPDQYRCTCPDGYSGRNCE-----KAEHACTSNPCANGGS 335
Qy      246 CHHVTG--ECSCPSGWMGTVC-----GQP-----CPEGFRGKNC- 277
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      336 CHEVPSGFECPCSGWSGPTCALDIDECASNPCAAGGTCVDQVDGFECICPEQWVGATCQ 395
Qy      278 --SQECQ---CHNGGTCAATG--QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCV 327
      :  |  :  |  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

Db 396 LDANECEGKPCLNAFSCKNLIGGYCDCIPGWKGINCHINVNDC-----RGQCQ 444
 Qy 328 NGGKCYH-VSG-ACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCHPMSGEC 385
 :|| | :| :| || | || | || | || | ||
 Db 445 HGGTCKDLVNGYQCVCPRGFGGRHCE-----LERDKC-ASSPCH-----SGG- 485
 Qy 386 ACKPGWSGLYCNETCSPGFYGEACQ---QICS---CQNGADCDSVTGK--CTCAPGFKGI 437
 |: | :|: | || | |: :| :||| |: | | | |
 Db 486 LCEDLADGFHCH--CPQGFSGPLCEVDVDLCEPSPCRNGARCYNLEGDYCACPDDFGGK 543
 Qy 438 DCST---PCPLGTYGINCSSRCGCKNDA-----VCSPVDGSCTCKAGWHGVDCS 483
 :|| ||| | | || :|| || | | : | : |
 Db 544 NCSVPREPCPGGA---CRVIDGCGSDAGPGMPGTAASGVCGP-HGRCVSQPGG---NFS 595
 Qy 484 IRCPSGTWGFGCN-----LTCQCLNGGAC-NTLDG-TCTCAPGWRGEKCE-----LP- 528
 | || | |: | | || | : :| | | || |: ||
 Db 596 CICS SGFTGT YCHENIDDC LGQPCRNGGT CIDEVDAFRFCFCPSGWEGELCDTNPNDC LDPD 655
 Qy 529 -----CQDGT YGLNCAER---CD---CSHADGCHPT--TGHCRCLPG 562
 | || | | | || ||: |: : | | | ||
 Db 656 PCHSRGRCYDLVNDFYCACDDGWKGKTCHSREFQCDAYTCSNGGT CYDSGDTFRACAPPG 715
 Qy 563 WSGVHCDSVCAEGRWGPNC SLPCYCKNGASC--SPDDGICECAPGFRGTTTCQRICSPGFY 620
 | | | | : :|| | || :| | | | : | ||
 Db 716 WKG---STCAVAK-NSSC-LPNPCVNGGTCVGS GASFS CICRDGWEGRTCT----- 761
 Qy 621 GHRC SQTC PQCVHSSGPCH HITGL---CDCLPGFTGALCN---EVCPSGRFGKNCAGIC 673
 | : | :| | : |: |:| || | | : | | ||
 Db 762 -HNTNDCNPLPCYNGGIC--VDGVNWFRCECAPGFAGPDCRINIDECQS----SPCAYGA 814
 Qy 674 TCTN--NGTCNPIDRSCQCYPGWIGSDC-----SQPCPPAHWGPNCIHTCNCHN 720
 || : || | || | | | | : | | | : : ||
 Db 815 TCVDEING-----YRCS CPPGRAGPRCQEVIGFGRSCWSRGTPFPH-GSSWVEDCNS-- 865
 Qy 721 GAFCSAYDGECKCTPGWTGLYCTQRCPLGFY GKDCALICQCQNGADC-DHISGQC---TC 776
 | || |: | | : | | |: || || | | : ||| |
 Db 866 ---CRCLDGRRDCSKVWCG---WKPCLLA--GQPEALSAQCPLGQRCLKAPGQCLRPPC 917
 Qy 777 RT-GFMGRHCEQKCP----SGTYGYGCRQICDCLNNSTCDHI----TGTCYCS-----PG 822
 | | | | || | :| | ||: | || | || |
 Db 918 EAWGECGAEPPSTPCLPRSGHLDNNCARLTLHFNR---DHVPQGT TVGAICSGIRSLPA 974
 Qy 823 WKGARCDQAGV IIVGNLNS-LSRTSTA-----LPADSYQIGAIAGI I I LVLVFL 873
 : |: |: :| | | || | || |: :
 Db 975 TRAVARDRLVL LCDRASSGASAVEVAVSFSPARDLPDSSLIQGA AHAIVAAI----- 1027
 Qy 874 ALFIIYRHKQKGKES SMPAVTYTPAMRVV 902
 |: | : || | ||
 Db 1028 -----TQRGNSSLL LAVTEVKVETVV 1048

Search completed: March 26, 2004, 16:13:05
 Job time : 46.759 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:05:25 ; Search time 23.117 Seconds
(without alignments)
4743.616 Million cell updates/sec

Title: US-10-092-390-2
Perfect score: 6744
Sequence: 1 MVISLNSCLSFICLLCHWI.....SSPKQEDSGGSSSNSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	1958	29.0	1574	2	T13954	MEGF6 protein - ra
2	1900	28.2	1620	2	T27283	hypothetical prote
3	1805.5	26.8	1111	2	T26972	hypothetical prote
4	1036	15.4	2524	2	A35844	Xotch protein - Af
5	1032	15.3	2555	2	A40043	notch protein homo
6	1028	15.2	2531	2	A46019	notch-1 protein -
7	1024	15.2	2531	2	S18188	notch protein homo
8	998	14.8	2471	2	A49128	cell-fate determin
9	987	14.6	2437	2	S42612	transmembrane prot
10	978.5	14.5	2703	1	A24420	notch protein - fr
11	977.5	14.5	2318	2	S45306	notch 3 protein -
12	974	14.4	2321	2	S78549	notch3 protein - h

13	964.5	14.3	2531	2	T31070	notch homolog - se
14	954.5	14.2	1064	2	A40136	fibropellin Ia - s
15	952.5	14.1	1964	2	T09059	notch4 - mouse
16	939.5	13.9	1203	2	A49175	Notch B protein -
17	920.5	13.6	4006	2	T09070	probable tenascin
18	915.5	13.6	2352	2	T30201	Notch homolog prot
19	914.5	13.6	3566	1	A40701	tenascin-X precurs
20	881.5	13.1	4135	2	T42629	tenascin-X - bovin
21	775.5	11.5	2139	2	A35672	crumbs protein - f
22	769	11.4	2201	2	A32160	tenascin-C - human
23	767.5	11.4	1746	1	S19694	tenascin precursor
24	746.5	11.1	2019	1	JQ1322	tenascin precursor
25	744	11.0	1220	2	A56136	jagged protein pre
26	720.5	10.7	861	2	A48825	Notch homolog Motc
27	717.5	10.6	1801	1	MMRTS	laminin beta-2 cha
28	716	10.6	3106	1	S53868	laminin alpha-2 ch
29	709.5	10.5	1810	1	A32230	tenascin precursor
30	704.5	10.4	3084	1	MMMSA	laminin alpha-1 ch
31	695.5	10.3	1790	1	MMFFB1	laminin beta-1 cha
32	692	10.3	3075	2	S14458	laminin alpha-1 ch
33	690	10.2	1408	2	S16148	gene serrate prote
34	685.5	10.2	1798	2	S53869	laminin beta-2 cha
35	683.5	10.1	3635	2	T10053	laminin alpha 5 ch
36	676.5	10.0	3672	2	T23433	hypothetical prote
37	676.5	10.0	3704	2	T37316	probable laminin a
38	671	9.9	1722	2	E89753	protein F11C7.4 [i
39	667	9.9	1786	1	MMHUB1	laminin beta-1 cha
40	666	9.9	2823	2	T23064	hypothetical prote
41	666	9.9	2823	2	F87908	protein T22A3.8 [i
42	666	9.9	3102	2	T43291	laminin alpha chai
43	664	9.8	1786	1	MMMSB1	laminin beta-1 cha
44	657	9.7	1429	2	S06434	homeotic protein 1
45	641	9.5	1797	2	A55677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T13954

MEGF6 protein - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C;Accession: T13954

R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.

A;Reference number: Z14126; MUID:98360089; PMID:9693030

A;Accession: T13954

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1574 <NAK>

A;Cross-references: EMBL:AB011532; NID:g3449293; PIDN:BAA32462.1; PID:g3449294

A;Experimental source: strain Sprague-Dawley; brain

C;Genetics:

A;Gene: MEGF6

Query Match 29.0%; Score 1958; DB 2; Length 1574;
 Best Local Similarity 41.3%; Pred. No. 2.6e-96;
 Matches 344; Conservative 77; Mismatches 306; Indels 106; Gaps 16;

```

Qy      95 CP-GFYESGEMCVP--HCADKCVHGRC-IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTS 150
      || ||| |: | |||:: ||| | :|| | | || :|| |
Db      602 CPKGFY--GKHCRKKCHCANR---GRCHRLYGACLCDPGLYGRFCHLACPPWAFGPGCSE 656

Qy      151 RCQCKNG--ALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTG 208
      | |: ||| |:| | |||:| |: || | :| | || | | || |:|
Db      657 DCLCEQSHTRSCNPKDGSCSKAGFQGERCQAECESGFFGPGCRHRCTCQPGVACDPVSG 716

Qy      209 ECR--CPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQG 266
      ||| |||| | | || | | | | | | | || |||| | | | |
Db      717 ECRTQCPPGYQGEDCGQECPVGTFGVNCSGSCSCV-GAPCHRVGTGECCLPPGKTGEDCGA 775

Qy      267 PCPEGRFGKNCSQEC-QCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQ 325
      ||||:| | : | | :| :|: || | | ||: | |||| | | || | |
Db      776 DCPEGRWGLGCQEICPACEHGASCNPETGTCLCLPGFVGSRCQDTCASAGWYGTGCQIRCA 835

Qy      326 CVNGG-----KCYH-----VSGACLC 341
      | | | | | | | | | | | | | | | |
Db      836 CANDGHCDPTTGRCSAPGWTGLSCQACDSGHWGPDCIHPNCNSAGHGNCDAVSGLCCLC 895

Qy      342 EAGFAGERCEARLCPEGLYGIKDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCS 401
      |||: | ||| : | :| || | :|| | | :| :|| | | | | :| |
Db      896 EAGYEGPRCE-QSCRQGYGSPCEQKRC--EHGAACDHVSGACTCPAGWRGSFCEHACP 952

Qy      402 PGFYG-----EA-----CQQICSCQNG 418
      ||:| :| | | | | | | | | | | |
Db      953 AGFFGLDCDSACNCSAGAPCDAVTGSCICPAGRWGPRCAQSCPPLTFGLNCSQICTCFNG 1012

Qy      419 ADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDVGSTCKAGWH 478
      | |||||:| ||||: | | || | || | | | :| | | :| | || |
Db      1013 ASCDSVTGQCHCAPGWMGPTCLQACPPGLYGKNCQHSLCRNGGRCDPILGQCTCPEGWT 1072

Qy      479 GVDCSI RCPSTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGINC 538
      |: | | | : || | | ||: || | : | | | | || :||: | ||:|:|
Db      1073 GLACENECLPGHYAAGCQLNCSCLHGGICDRLTGHCCLCPAGWTGDKCQSSCVSGTFGVHC 1132

Qy      539 AERCDCSHADGCHPTTGHCRCLPGWSGVHCDVCAEGRWGPNCSLPCYCKNGASCSPDDG 598
      | | | | | | | | | | | | :| | | :| | | | | |
Db      1133 EEHCACRKGASCHHVTGACFCPPGWRGPHCEQACPRGWFGGEACAQRCLCPTNASCHHVTG 1192

Qy      599 ICECAPGFRGTTCQRICSPGFYGHRCSTCPQC CVHSSGPCHHITGLCDCLPGFTGALCNE 658
      | | ||| | :||: | || :| | || | : | :|| | | :| | | :
Db      1193 ECRCPPGFTGLSCEQACQPGTFGKDCEHLC-QCPGETWACDPASGVCTCAAGYHGTGCLQ 1251

Qy      659 VCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNC 718
      ||||:| | || | | |||:| :| | ||:|:| || | :||:| | | |
Db      1252 RCPSGRYGPGCEHICKLNGGTCDPATGACYCPAGFLGADCSLACPQGRFGPSCAHVCAC 1311

Qy      719 HNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKGDCALICQCQNGADCDHISGQCTCRT 778
      || | | | :|| ||: | || :|| | | | :|| | | :| | :|
Db      1312 RQGAACDPVSGACICSPGKTGVRCEHGCPQDRFGKGCELKCACRNGGLCHATNGSCSCPL 1371

```

Qy 779 GFMGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQA 831
|:| | | | | :| | | | | | | :| | | | | :| | :
Db 1372 GWMGPHCEHACPAGRYGAACLLECFQNNGSCEPTTGACLCGPGFYGQACEHS 1424

RESULT 2

T27283

hypothetical protein Y64G10A.f - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T27283

R;Ainscough, R.

submitted to the EMBL Data Library, September 1999

A;Reference number: Z20336

A;Accession: T27283

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1620 <WIL>

A;Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB54471.1; CESP:Y64G10A.f

A;Experimental source: clone Y64G10A

C;Genetics:

A;Gene: CESP:Y64G10A.f

A;Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1;

625/1; 715/1; 782/3; 845/1; 895/2; 956/1; 1105/1; 1221/1; 1307/1; 1445/2

Query Match 28.2%; Score 1900; DB 2; Length 1620;

Best Local Similarity 38.6%; Pred. No. 3.1e-93;

Matches 331; Conservative 88; Mismatches 307; Indels 132; Gaps 17;

Qy 95 CP-GFYESGEMCVPHCADKCVHGRCIAP-NTCQCEPGWGGTNCSSACDGDHWGPHCTSRC 152
| | | | | | | | | | :| | | | | | | :| | | | |
Db 712 CPDGFY--GSQCNLKRMDCPNGRCDPVFGYCTCPDGLYGQSCEKPCPHFTFGKNCRFPC 769

Qy 153 QC--KNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGEC 210
:| :| | :| | | | | :| | | :| | | | | | | :|
Db 770 KCARENSEGCD EITGKCRCKPGYYGHHCKRMCSPLFGAGCAMEKSCSPAGIRCDPVTGDC 829

Qy 211 --RCPPGYTGAFCEDLCP PGKHGPQCEQRCPC-----QNGGVCHHVTGECSCPSGWMGT 262
:| | | | | :| | | | | | | | | | | | | | :| | | |
Db 830 TKKCPAGYQGNLCDQPCPAGYFGYDCEQKCSADVASPHKSKVCHHVTGTCTCLPGKTGP 889

Qy 263 VCGQ-----PCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYT 303
:| | | | | | | | | | :| | | | | | | | :| | | | :|
Db 890 LCDQCLIFVETIEFDIAFSINVIACAPNTYGPNC AHTCSCVNGAKCDESDGSCHCTPGFY 949

Qy 304 GERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIK 363
| | :| | | :| :| :| | | | | :| :| | | :| :| :| :| :| :| :|
Db 950 GATCSEVCPTGRFGIDCMQLCKCQNGAICDTSNGSCECAPGWSGKKCD-KACAPGTFGKD 1008

Qy 364 CDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCD 423
| :| | | :| | | | | | | | | | :| | | :| | | | | | | | |
Db 1009 CSKKCDC-ADGMH-CDPSDGEICPPGKKGHKCD ETCDSGLFGAGCKGICSCQNGATCDS 1066

Qy 424 VTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGC--KND-----AVCSPVDGSCCTC 473
| | | | | | :| | | | | :| :| | | | | | | | | | | |
Db 1067 VTGSCECRPGWRGKKCDRPCPDGRFGE GCNAICDCTTTNDTSMYNPFVARCDHVTGECRC 1126

Qy 474 KAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQD-- 531
 ||| | || | | | | :||| | :| : | | | | :| : | | | :
 Db 1127 PAGWTGPDCQTSCPLGRHGEGCRHSCQCSNGASCDRVTGFCDCPSGFMGKNCESECPEGL 1186

Qy 532 -----GTYGLNCAERCDCSHADGCHPT 553
 | :| |||:|:| : |
 Db 1187 WGSNCMKHCLCMHGGECNKENGDCCEIDGWTGPSLCPFGQFGRNCAQRCNCKNGASCDRK 1246

Qy 554 TGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTTTCQR 613
 || | ||||| ||: | | :| | | :||| | | | | ||:| | |
 Db 1247 TGRCECLPGWSGEHCEKSCVSGHYGAKCEETCECENGALCDPISGHCSQCPGWRGKKCNR 1306

Qy 614 ICSPGFYGHRSQTCPCQCVHSSGPPCHHITGLCDCLPGFTGALCNEVCP----- 661
 | | :| | ||:| :| :| | ||:| | | | :| | | :||
 Db 1307 PCLKGYFGRHCSQSC-RCANSKS-CDHISGRCQCPKGYAGHSCTELCPDGTFGESCSQKC 1364

Qy 662 -----SGRFGKNCAGICTCTNNGTCNPIDRSCQC 690
 |||| :| :|:| | | | :| || |
 Db 1365 DCGENSMCDALSGKCFCKPGHSGSDCKSGCVQGRFGPCDNQLCSCENGGVCDSSSTGSCVC 1424

Qy 691 YPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGF 750
 ||:|:| | | :|| | ||| || | | :|:| ||:|:| | | || |
 Db 1425 PPGYIGTKCEIACQSDRFGPTCEKICNCENGGTCDRLTGQCRCPLPGFTGMTCNQVCPEGR 1484

Qy 751 YGKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNSTC 810
 :| | | :| || | : ||:| | || | ||| |||| | | | :| : |
 Db 1485 FGAGCKEKCRCANG-HCNASSGECKCNLGFTGPSCEQSCPSGKYGLNCTLDCECYGQARC 1543

Qy 811 DHITGTCYCSPGWKGARC 828
 | : | | | || | :||
 Db 1544 DPVQGCCDCPPGRYGSRC 1561

RESULT 3

T26972

hypothetical protein Y47H9C.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C;Accession: T26972

R;Harris, B.

submitted to the EMBL Data Library, October 1998

A;Reference number: Z20293

A;Accession: T26972

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1111 <WIL>

A;Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4

A;Experimental source: clone Y47H9C

C;Genetics:

A;Gene: CESP:Y47H9C.4

A;Map position: 1

A;Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1

C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 26.8%; Score 1805.5; DB 2; Length 1111;

Best Local Similarity 31.8%; Pred. No. 2.3e-88;

Matches 373; Conservative 162; Mismatches 423; Indels 215; Gaps 40;

Qy 21 GTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYT-----SCTDILNWFKCTRHR 73
|| | : :||: | : : | : : | : | :|
Db 35 GTTEP---QGDHVCT----VKTIVDDY--ELKKVIHTVVYNDTEQCLNPLTGFQC---- 80

Qy 74 VSYRTAYRHGEKTMYYRK-----SQCCPGFYESGE-MCVPHCADKCVHGRCIAPNTC 124
| : :| | :| : || | :| | | :|| | |
Db 81 ----TVEKRGQKASYQRQLVKKEKYVKQCCDGYQTKDHFCLPDCNPPCKKGKCIEPGKC 136

Qy 125 QCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQ 184
:|:|:| | :|:| | | :| :| | | :|:| | | |
Db 137 ECDPGYGGKYCASSCSVGTWGLGCSKSCDCENGANCDELGTCTCTSGFQGERCEKPCPD 196

Qy 185 GTYGNDCHQRCQCCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGG 244
:| :| :| | | | :| :| | | :| | | :| :| :| | | |
Db 197 NKWGPNCVKSCPCQNGGKCNK-EGKCVCSDBGWGEFCLNKCEEKFGAECKFECNCQNGA 255

Qy 245 VCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTG 304
| : | :| | | :| :| | | | :| :| | | | :| :| :| |
Db 256 TCDNTNGKCICKSGYHGALCENECVSGFFGSGCTQKCDCLNNQNCSSSGECKCI-GWTG 314

Qy 305 ERCQDECPVGTYGVLCAETCQCV-----NGGKCYHVSAGLCEAGFAGERCEARLCPEG 358
: | | | :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 315 KHCDIGCSRGFRGLQCKQNCCTCPGLEFSDSNASCDAKTGQCQCESGYKGPCKDERKCDAE 374

Qy 359 LYGIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEAC--QQICSCQ 416
|| | | | | | | | :| | | | :| | | | | | :| :|
Db 375 QYGADCSKTCTCVRENTLMCAPNTGFCRCRKPFGYGDNCCELACSKDSYGPNCQKAMCDWN 434

Qy 417 NGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVAV-CSPVDGSGCTCKA 475
: :|:| | | | | :| | | | | | :| :| | | | | | |
Db 435 HASECNPETGSCVCKPGRTGKNCSEPCPLDFYGPNCQAHQCCNQRGVGCDGADGKCQCDR 494

Qy 476 GWHGVDCSIRCPSGTWGFGCNLTQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYG 535
|| | | | :| :| | | :| | | :| :| | | | :| :| :| :|
Db 495 GWTGHRCEHHCPADTFGANCKRCKCPKGIGCDPITGECTCPAGLQGANCDIGCPEGSYG 554

Qy 536 LNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSP 595
| | | :| | | | | :| | | :| :| :| :| :| | | :| | |
Db 555 PGCKLHCKCVNGK-CDKETGECTCQPGFFGSDCSTTCSKGKYGESCELSCPCSD-ASCSK 612

Qy 596 DDGICECAPGFRGTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHI---TGLC-DCLPGF 651
| | | | :| :| :| | :| | :| :| | | :| :| | |
Db 613 QTGKCLCPLGTGKGVSCDQKCDPNTFGFLCQETV-----TPSPCASTDPKNGVCLSCPPGS 667

Qy 652 TGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGN 711
:| | | | :| :| :| :| :| :| | | | :| :| :| :|
Db 668 SGIHCEHNCPAGSYGDGCQQVCSCADGHGCDPTTGECICEPGYHGKTCSEKCPDGKYGYG 727

Qy 712 CIHTC-NCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCCQNG-ADCDH 769
| | | :| :| :| | | | | :| | | :| :| :| :| :|
Db 728 CALDCPKCASGSTCDHINGLICIPAGLEGALCTRPCSAGFWGNGCRQVCRCTSEYKQCNA 787

Qy 770 ISGQCTCRTGFMGRHCEQKCPSTGYGYGCRQICDCLNNST--CDHITGTCYCSPGWKGAR 827
:|:|:| | | | :| :| | | | :| :| :| :| :| :| :|
Db 788 QTGECSCPAGFQGDRCDKPCEDGYYGPDICIKKCKCQGTATSSCNRVSGACHCHPGFTGEF 847

```

Qy      828 CDQAGVIIVGNLNSLSRTST-----ALP----ADSYQIGAIAG----- 861
      |          ::|  ||          |  | :  |  |
Db      848 C-----HALCPESTFGLKCSKECPKDGCGDGYECDAAIGCCHVDQMSCGKAKQE 896

Qy      862 -----IIILVLVVLVLF----LLALFIIYRHK-QKGKESMPAVTYTPAMRV 901
      |:::|  |  |::|  ||:|  ||  |:  ||  ::
Db      897 FEALNGAGRSTGLTWFFVLLIVALCGGLGLIALF--YRNKYQKEKDPDMPTVSF----- 948

Qy      902 VNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVN 961
      |  |  |  ||  |  ::  :  |  :  |  :  |:|
Db      949 -----HKAPNNDEGREFQNPPLY---SRQSVFP---DSDAFSSENNGNHQ---- 986

Qy      962 LKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGLDRSYM-----GKSLKDLGKN-- 1013
      ||  |  |  |:  :  |:|  |  |
Db      987 -----GGPPN--GLLTLEEEELLENKKIHGRSAAGRGNNDY 1019

Qy      1014 -----SEYNSSNCSLSSE---NPYATIKDPPVLIPKSSECGYVEMK----SPA---- 1055
      | :||:  | :|  |  :  |  |:  :  |  :  |  ||
Db      1020 ASLDEVAGEGSSSSASASASRRGLNSSEQSRRP--LLEEHDEEEFDEPHENSISPAAVAVT 1077

Qy      1056 ---RRDSPYAEINN-----STSANR-----NVY 1075
      ::|||:|:  ||:  |:|
Db      1078 TSNHNENPYADISSPDPVTQNSANKKRAQDNLY 1110

```

RESULT 4

A35844

Xotch protein - African clawed frog

C;Species: *Xenopus laevis* (African clawed frog)

C;Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 02-Aug-2002

C;Accession: A35844

R;Coffman, C.; Harris, W.; Kintner, C.

Science 249, 1438-1441, 1990

A;Title: Xotch, the *Xenopus* homolog of *Drosophila* notch.

A;Reference number: A35844; MUID:90385285; PMID:2402639

A;Accession: A35844

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-2524 <COF>

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

C;Keywords: transmembrane protein

F;146-177/Domain: EGF homology <EGX1>

F;184-215/Domain: EGF homology <EGF1>

F;222-254/Domain: EGF homology <EGF>

F;456-487/Domain: EGF homology <EGX2>

F;757-788/Domain: EGF homology <EGF3>

F;1025-1056/Domain: EGF homology <EGX3>

F;1924-1956/Domain: ankyrin repeat homology <AN1>

F;1957-1989/Domain: ankyrin repeat homology <AN2>

F;1991-2023/Domain: ankyrin repeat homology <AN3>

F;2024-2056/Domain: ankyrin repeat homology <AN4>

F;2057-2089/Domain: ankyrin repeat homology <AN5>

Query Match 15.4%; Score 1036; DB 2; Length 2524;

Best Local Similarity 25.4%; Pred. No. 2.6e-47;

Qy 565 GVHCD-----SVCAEGRWGPNC SL-----PCYCKNGASC 593
 | | : | | | | : | | | |
 Db 863 GQTCEIDMNECVNRPCRNGATCQNTNGSYKCNCKPGYTGRNCEMDIDDCQPNPCHNGGSC 922

 Qy 594 SPDDGI----CECAPGFRGTTTCQR-----ICSPGFYGHRC 624
 | | | | | | | : | | | | |
 Db 923 S--DGINMFFCNCNPAGFRGPKCEEDINECASNPCKNGANCTDCVNSYTCTCQPGFSGIHC 980

 Qy 625 SQTCPQCVHSS----GPCHHITGL---CDCLPGFTGALC----NE----- 658
 | | | | | | : | | | | | : | |
 Db 981 ESNTPDCTESSCFNGGTC--IDGINTFTCQCPPGFTGSYCQHDINECDSPCLNGGTCQD 1038

 Qy 659 -----VCPSGRFPGKNCAGI-----CTCTNNGTC--NPIDRSCQCYPGWIGSDCSQP 702
 | | | | | : | | | | | : | | | | |
 Db 1039 SYGTYKCTCPQGYTGLNCQNLVWRWCDSSPCKNGGKCWQTNNFYR-CECKSGWTGVYCDVP 1097

 Qy 703 ---CPPA--HWGPNCIHTCNCHNGAFC--SAYDGECKCTPGWTGLYCTQRCPLGFYKDC 755
 | | : : | | : : | : | | : : : |
 Db 1098 SVSCEVAAKQQGVDIVHL--CRNSGMCVDTGNTHFCRCQAGYTGSYCEEQV-----DEC 1149

 Qy 756 ALICQCQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 : | | | | | : : | | : : : :
 Db 1150 S-PNPCQNGATCTDYLGGSCECVAGYHGVNCSSEEINECLSHPCQNGGTCIDLINTYKCS 1208

 Qy 789 CPSGTYGYGCRQICD-----CLNNSTC-DHITG-TCYCSPGWKGARCDQAG 832
 | | | | | | | : | | | | : | | :
 Db 1209 CPRGTQGVHCEINVDDCTPFYDSFTLEPKCFNNGKCIDRVGGYNCICPPGFVGERCE--- 1265

 Qy 833 VIIVGNLNS-LSRTSTALPADS 853
 | : : | | | |
 Db 1266 ----GDVNECLSN-----PCDS 1278

RESULT 5

A40043

notch protein homolog TAN-1 precursor - human

C;Species: Homo sapiens (man)

C;Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 02-Aug-2002

C;Accession: A40043

R;Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, J.

Cell 66, 649-661, 1991

A;Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms.

A;Reference number: A40043; MUID:91347367; PMID:1831692

A;Accession: A40043

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-2555 <ELL>

A;Cross-references: GB:M73980

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;261-292/Domain: EGF homology <EGX1>

F;494-525/Domain: EGF homology <EGF1>

F;987-1018/Domain: EGF homology <EGX2>

F;1149-1180/Domain: EGF homology <EGF>

F;1187-1218/Domain: EGF homology <EGF3>
 F;1233-1264/Domain: EGF homology <EGX3>
 F;1927-1959/Domain: ankyrin repeat homology <AN1>
 F;1960-1992/Domain: ankyrin repeat homology <AN2>
 F;1994-2026/Domain: ankyrin repeat homology <AN3>
 F;2027-2059/Domain: ankyrin repeat homology <AN4>
 F;2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 15.3%; Score 1032; DB 2; Length 2555;
 Best Local Similarity 25.7%; Pred. No. 4.2e-47;
 Matches 315; Conservative 84; Mismatches 304; Indels 522; Gaps 73;

```

Qy      94 CCPGFYESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGGTNCSSACDGDH 143
      | | | | : | : | : | : | : | : | : | : |
Db      89 CALGF--SGPLCLTPLDNACLTNPCRNGGTCDLLTLTEYKCRCPPGWSGKSCQQA----- 141

Qy     144 WGPHTSRCQCKNGALCNPITGA--CHCAAGFRGWRCE---DRCEQG---TYGNDCHQ- 193
      | | | | | | : | | | | | : | | : | | |
Db     142 --DPCASN-PCANGGQCLPFEASYICHCPFSFHGPTCRQDVNECGQKPRLCRHGGTCHNE 198

Qy     194 ----RC-----QCQNGATC---DHVTGECRCPPGYTGAFCE--- 222
      | | | | | | | | | | | | | | | |
Db     199 VGSYRCVCRATHTGPNCEPYVPCSPSPCQNGGTCRPTGDTVTHEACALPGFTGQNCENI 258

Qy     223 DLCPPG--KHGPQC-----EQRCP-----CQNGGVCHHVTG- 251
      | | | | : | | | | | | | | | | : | |
Db     259 DDCPGNNCKNGGACVDGVNTYNCPCPPEWTGQYCTEDVDECQLMPNACQNGGTCKNTHGG 318

Qy     252 -ECSCPSGWMGTVCQG-----PCPEGRFGKNC--SQEC-- 281
      | | : | | | : | | | | | | : | |
Db     319 YNCVCVNGWTGEDCSENIDDCASAACFHGATCHDRVASFYCECPHGRTGLLCHLNDACIS 378

Qy     282 -QCHNGGTCDA--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCATCQCVNGGKCY 333
      | : | | | : | | | | | | | | : | | : | |
Db     379 NPCNEGSNCDTNPVNGKAICTCPSGYTGPAQSQDVDECSLGAN-----PCEHAGKCI 430

Qy     334 HVSGA--CLCEAGFAGERCEARLCPEGLYGIKCDKRC---PCHLENTHSCHPMSE--CA 386
      : | : | | : | | | : | | | : | : | | |
Db     431 NTLGSFECQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477

Qy     387 CKPGWSGLYC-----NE---TCSPGFYGEACQOI--CS---CQ 416
      | | : | : | | | | | | | | : | : | : |
Db     478 CMPGYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCQDVDECASTPCK 537

Qy     417 NGADC-----DSV-TGKCTCAPGFKGI 437
      | | | | | | | | | | | | : | |
Db     538 NGAKCLDGPNTYTCVCTEGYTGTHTCEVDIDECDPDPCHYGSKDGVATFTCLCRPGYTGH 597

Qy     438 DCST-----PCPL-----GTYGINCS-----SRCGCKNDAVCSP 466
      | | | | | | | | | | : | : |
Db     598 HCETNINECSSQPCRLRGTCQDPDNAYLCFCLKGTTGPNCEINLDDCASSPCDSGTCLDK 657

Qy     467 VDG-SCTCKAGWHGVDCSIR-----CPSGTWGFNCNL-----TC----- 499
      : | | | : | : | : | | | | | | |
Db     658 IDGYECACEPGYTGSMCNSNIDECAGNPCHNGGTCEDGINGFTCRCPEGYHDPTCLSEVN 717

Qy     500 -----QCLNGGACNTLDG-TCTCAPGWRGEKCEL-----P 528

```

```

      |::|   ::::| | | | | | |::
Db      718 ECNSNPCVHGACRDSLNGYKCDGDPGWSGTNCDINNNECESNPCVNGGTCKDMTSGIVCT 777

Qy      529 CQDGTYGLNCAERCD-----CSHADGC-HPTTGH-CRCLPGWSGVHCDV---CAEG-- 575
      |::| | | | : | | | | | |:: | | | |
Db      778 CREGFSGPNCQTNINECASNPCLNKGTCIDDVAGYKCNCLLPYTGATCEVVLAPCAPSPC 837

Qy      576 RWGPNC-----SLPCYC----KNGASCSPDDGICECAPGFRGTTTCQRI-----CSP 617
      | | | | | | | | | | | | | | | |
Db      838 RNGGECRQSEDIYESFSCVCPTAGAKGTCEVDINECVLSPCRHGASCQNTHGXYRCHCQA 897

Qy      618 GFYGHRCSTQCPQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
      | : | | | | | | | | | | | | | | |
Db      898 GYSGRNCETDIDDC--RPNPCHNGGSCTDGINTAFCDCLPGFRGTFCCEEDINECASDPCR 955

Qy      659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCPYPGW 694
      | | | | | | | | | | | | | | | |
Db      956 NGANCTDCVDSYTCTCPAGFSGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGF 1013

Qy      695 IGSDC-----SQP-----CPPAHWGPNC---IHTCN---CHNGAF 723
      | | | | | | | | | | | | | | | |
Db      1014 TGSYCQHVVNECDSRPCLLGGTCQDGRGLHRCTCPQGYTGPNCQNLVHWCDSSPCKNGGK 1073

Qy      724 C----SAYDGECKCTPGWTGLYCTQ----- 744
      | : | | | | | | | | | | | | | |
Db      1074 CWQHTQY--RCECPSGWTGLYCDVPSVSCEVAAQRQGVVVARLCQHGGLCVDAGNTHHC 1131

Qy      745 RCPLGFYGKDCALI---CQ---CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
      | | | : | | | : | | | | | | | | | | |
Db      1132 RCQAGYTGSYCEDLVDECSPPSPCQNGATCTDYLGGYSCKCVAGYHGVNCSEEIDECLSHP 1191

Qy      789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG- 815
      | | | | | | | | | | | | | | | |
Db      1192 CQNGGTCLDLPLNTYKCSCPRGTTQGVHCEINVDDCNPPVDPVSRSPKCFNNGTCVDQVGGY 1251

Qy      816 TCYCSPGWKGARCDQAGVIIIVGNLN 840
      : | | | | : | | | | | : | |
Db      1252 SCTCPPGFVGERCE-----GDVN 1269

```

RESULT 6

A46019

notch-1 protein - mouse

N;Alternate names: motch protein

C;Species: Mus musculus (house mouse)

C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Mar-2003

C;Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109

R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Gridley, T.

Genomics 15, 259-264, 1993

A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of Drosophila Notch.

A;Reference number: A46019; MUID:93194170; PMID:8449489

A;Accession: A46019

A;Status: not compared with conceptual translation

A;Molecule type: nucleic acid

A;Residues: 1-2531

A;Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
 A;Note: sequence extracted from NCBI backbone (NCBIP:127318)
 R;Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.; McMahon, A.P.; Gridley, T.
 submitted to the EMBL Data Library, April 1992
 A;Description: Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development.
 A;Reference number: S25144
 A;Accession: S25144
 A;Molecule type: mRNA
 A;Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
 A;Cross-references: EMBL:Z11886
 R;Lardelli, M.; Lendahl, U.
 Exp. Cell Res. 204, 364-372, 1993
 A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues.
 A;Reference number: A49175; MUID:93178563; PMID:8440332
 A;Accession: C49175
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1161-1547 <LAR>
 A;Cross-references: EMBL:X68278; NID:g287987; PIDN:CAA48339.1; PID:g287988
 A;Experimental source: embryo
 A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
 R;Kopan, R.; Weintraub, H.
 J. Cell Biol. 121, 631-641, 1993
 A;Title: Mouse notch: expression in hair follicles correlates with cell fate determination.
 A;Reference number: A46438; MUID:93252998; PMID:8486742
 A;Accession: B46438
 A;Status: preliminary
 A;Molecule type: nucleic acid
 A;Residues: 1865-1932, 'RR', 1935-1937, 'L', 1938-1967, 'I', 1969-2044, 'IE', 2047-2052, 'S', 2054-2056, 'SIRRE', 2062-2075 <KOP>
 A;Experimental source: embryo
 A;Note: sequence extracted from NCBI backbone (NCBIN:131246, NCBIP:131247)
 C;Comment: This protein has many EGF repeats and lin-12[#1172]/Notch repeats.
 C;Comment: This protein is one of the neurogenic proteins controlling the decision between ectodermal and neural fate for cells in the early embryo.
 C;Genetics:
 A;Gene: notch-1
 A;Map position: 2
 A;Note: proximal region of chromosome 2
 C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F;106-138/Domain: EGF homology <EGF1>
 F;144-175/Domain: EGF homology <EGF1>
 F;222-254/Domain: EGF homology <EGF2>
 F;261-292/Domain: EGF homology <EGF2>
 F;339-370/Domain: EGF homology <EGF3>
 F;416-449/Domain: EGF homology <EGF3>
 F;456-487/Domain: EGF homology <EGF3>
 F;494-525/Domain: EGF homology <EGF3>
 F;532-563/Domain: EGF homology <EGF3>
 F;607-638/Domain: EGF homology <EGF3>
 F;682-713/Domain: EGF homology <EGF3>
 F;757-788/Domain: EGF homology <EGF3>

F;795-826/Domain: EGF homology <EG10>
 F;873-904/Domain: EGF homology <EG11>
 F;911-942/Domain: EGF homology <EG12>
 F;949-980/Domain: EGF homology <EG13>
 F;987-1018/Domain: EGF homology <EG14>
 F;1025-1056/Domain: EGF homology <EG15>
 F;1063-1094/Domain: EGF homology <EG16>
 F;1149-1180/Domain: EGF homology <EG17>
 F;1187-1218/Domain: EGF homology <EG18>
 F;1233-1264/Domain: EGF homology <EGF4>
 F;1352-1383/Domain: EGF homology <EG19>
 F;1391-1425/Domain: EGF homology <EGF>
 F;1917-1948/Domain: ankyrin repeat homology <AN1>
 F;1949-1981/Domain: ankyrin repeat homology <AN2>
 F;1983-2015/Domain: ankyrin repeat homology <AN3>
 F;2016-2048/Domain: ankyrin repeat homology <AN4>
 F;2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 15.2%; Score 1028; DB 2; Length 2531;
 Best Local Similarity 25.7%; Pred. No. 6.8e-47;
 Matches 314; Conservative 83; Mismatches 286; Indels 538; Gaps 73;

Qy	86	TMYYRRKSQCCPGFYESGEMCV---PHCADKCVH-GRCI---APNTCQCEPGWGGTNCSS-	137
		: : : : : : : : : : : : : :	
Db	121	TLTEYKCRCS PGW--SGKSCQQADPCASNPCANGGQCLPFESSYICRCP PGFHGPTCRQD	178
Qy	138	-----ACDGDHWGPHC-----TSRCQCKNGALCNP----	162
Db	179	VNECSQNPGLCRHGGHCHNEIGSYRCACCATHTGPHCEL PYVPCSPSPCQNGATCRPTGD	238
Qy	163	ITGACHCAAGFRGWRCEDRCEQGT YGNDCHQRCQ CQNGATC-DHV-TGECRCPPGYTGAF	220
		: : : : : : : : : :	
Db	239	TTHECACLP GFAGQNCEENV D-----DCPGN-NCKNGGACVDGVNTYNCRCPP EVTGQY	291
Qy	221	C-EDLCPPGKHGPQCEQRC P--CQNGGVCHHVTG--ECSCPSGWMGTVCQG-----	266
		: : : : : : : : : : : : :	
Db	292	CTEDV-----DEC-QLMPNACQNA GTCHNTHGGYNCVCVNGWTGEDCSENIDDCASAA	343
Qy	267	-----PCPEGRFGKNC--SQEC--QCHNGGTCD A--ATGQ--CHCSPG	301
		: :	
Db	344	CFQGATCHDRVASFYCECPHGRTGLLCHLKHACISNPCNEGSNCDTNPVNGKRICTCPSG	403
Qy	302	YTGERCQ---DECPVGT YGVLCAETCQCVNGGKCYHVSGA--CLCEAGFAGERCEARLCP	356
		: : : : : : : : : :	
Db	404	YTGPACSQDVDECD LGAN-----RCEHAGKCLNTLG SFECQCLQGYTGPGCEIDV--	453
Qy	357	EGLYGIKCDKRC---PCHLENTHSCHPMSGE--CACKPGWSGLYCNET-----	399
		: : : :	
Db	454	-----NECISNPC--QNDATCLDQIGEFQCICMPGYEGVYCEINTDECASSPCLHN	502
Qy	400	-----CSPGFYGEACQ---QICS--CQNGADC--DSVTGKCTCAPGFKGID	438
		: : :	
Db	503	GHCMDKIHEFQCQCPKGFNGHLCQYDVDECASTPCKNGAKCLDGPNTYTCVCTEGYTGTH	562
Qy	439	CST-----PCPLGT-----YGINCSSRCG-----CKNDAVCSPVD	468
		: : : : : : :	
Db	563	CEVDIDECDPDPCHYGSKDGVATFTCLCQPGYTGHHCETNINECHSQPCRHHGGTCQDRD	622

Qy 469 GS--CTCKAGWHGVDCSIR-----CPSGTWGFGCNLTQCQLNGGACNTLDG-TCTCA 517
 | | | | : | | | | | : : | | |
 Db 623 NSYLCLCLKGTGPNCEINLDDCASNP CDSGT-----CL-----DKIDGYECACE 667

Qy 518 PGWRGEKCEL-----PCQDGTGYGLNCAERCDCSHADGCHPTT----- 554
 ||: | | : | : || | | | | : | | |
 Db 668 PGYTGSMCNVNIDEACAGSPCHNGGTCELDGIAGFTC--RC---PEGYHDPTCLSEVNECN 721

Qy 555 -----GHCN-----CLPGWSGVHCD-----SVCAE 574
 | | | | | | | | : | | | | |
 Db 722 SNPCIHGACRDGLNGYKCDCA PGWSGTNCDINNECESNPCVNGGTCKDMTSGYVCTCRE 781

Qy 575 GRWGPNC-----SLPCY-----CKNG 590
 | | | | | | | | | | | | | |
 Db 782 GFSGPNCQTNINECASNPCLNQGT CIDDVAGYKCNCLPYTGATCEVVLAPCATSPCKNS 841

Qy 591 ASCSPDDGI----CECAPGFRGTTTCQ-----RICSPGFYG 621
 | : | | | : | | | : | | : |
 Db 842 GVCKESEDYESFSCVCPTGWQGTCEVDINECVKSPCRHGASCQNTNGSYRCLCQAGYTG 901

Qy 622 HRCSQTCPCQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
 | | | | | | | | | | | | | |
 Db 902 RNCESDIDDC--RPNPCHNGGSCTDGINTAFCDCLPGFQGAFCCEEDINECASNPCQNGAN 959

Qy 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPGWIGSD 698
 | | | | : | | | | | : | | | : | |
 Db 960 CTDCVDSYTCCTCPVGFNGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGFTGSY 1017

Qy 699 C-----SQP-----CPPAHWGPNC---IHTCN---CHNGAFC--- 724
 | : | | | : | | | : | : | | | |
 Db 1018 CQYDVNECDSRPCLHGGTCQDSYGYKCTCPQGYTGLNCQNLVRWCDSAPCKNGGRCWQT 1077

Qy 725 -SAYDGECKCTPGWTGLYC---TQRCPLGFY--GKDCALICQ----- 760
 : | | : | | | : | : | | | : | |
 Db 1078 NTQY--HCECRSGWTGVNCDVLSVSCEVAAQKRIGDVTLLCQHGGLCVDEGDKHYHCQA 1135

Qy 761 -----CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 | | | | | | : | | | : | : | :
 Db 1136 GYTGSYCEDEVDECSNPNCQNGATCTDYLGGFSCCKCVAGYHGSNCSEEINECLSQPCQNG 1195

Qy 789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG-TCYC 819
 | | | | | | | | | | | | : | | | |
 Db 1196 GTCIDLTSYKCS CPRGTQGVHCEINVDDCHPPLDPASRSPKCFNNGTCVDQVGGYTCTC 1255

Qy 820 SPGWKGARCDQAGVIIIVGNLN 840
 ||: | | : | : |
 Db 1256 PPGFVGERCE-----GDVN 1269

RESULT 7

S18188

notch protein homolog - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002

C;Accession: S18188

R;Weinmaster, G.; Roberts, V.J.; Lemke, G.

Development 113, 199-205, 1991

A;Title: A homolog of Drosophila Notch expressed during mammalian development.

A;Reference number: S18188; MUID:92111383; PMID:1764995

A;Accession: S18188

A;Molecule type: mRNA

A;Residues: 1-2531 <WEI>

A;Cross-references: EMBL:X57405; NID:g57634; PID:g57635

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;987-1018/Domain: EGF homology <EGF1>

F;1025-1056/Domain: EGF homology <EGF>

F;1233-1264/Domain: EGF homology <EGF2>

F;1917-1949/Domain: ankyrin repeat homology <AN1>

F;1950-1982/Domain: ankyrin repeat homology <AN2>

F;1984-2016/Domain: ankyrin repeat homology <AN3>

F;2017-2049/Domain: ankyrin repeat homology <AN4>

F;2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 15.2%; Score 1024; DB 2; Length 2531;

Best Local Similarity 25.8%; Pred. No. 1.1e-46;

Matches 315; Conservative 83; Mismatches 286; Indels 536; Gaps 74;

```
Qy      86 TMYRRKSQCCPGFYESGEMCV---PHCADKCVH-GRCI---APNTCQCEPGWGGTNCSS- 137
      | : | : | | : | | : | : | : | : | : | : | : | : | : | : |
Db     121 TLTEYKCRCPPGW--SGKSCQQADPCASNPCANGGQCLPFESSYICGCPPGFHGPCTCRQD 178

Qy     138 -----ACDGDHWGPHC-----TSRCQCKNGALCNP--- 162
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     179 VNECSQNPLCRHGGTCHNEIGSYRCACRATHTGPHCELPHYVPCSPSPCQNGGTCRPTGD 238

Qy     163 ITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATC-DHV-TGECRCPPGYTGAF 220
      | | | | | | | | : | : | | | | | | | | | | | | | : | : |
Db     239 TTHECACLPGFAGQNCENVD-----DCPGN-NCKNGGACVDGVNTYNCRCPPPEWTGQY 291

Qy     221 C-EDLCPPGKHGPQCEQRCP--CQNGGVCHHVTG--ECSCPSGWMGTVCQG----- 266
      | | | : | : | | | | | | | | : | | | | | | | | | | | |
Db     292 CTEDV-----DEC-QLMPNACQNAQTCHNSHGGYNCVVCVNGWTGEDCSDNIDDCASAA 343

Qy     267 -----PCPEGRFPGKNC--SQEC---QCHNGGTCD--ATGQ--CHCSPG 301
      | | | | | | : | | | | | | | | | | | | | | | | | | |
Db     344 CFQGATCHDRVASFYCECPHGRTGLLCHLNDACISNPCNEGSNCDTNPVNGKAICTCPRG 403

Qy     302 YTGERCQ---DECPVGTYGVLCAETCQCVNGGKCYHVSGA--CLCEAGFAGERCEARLCP 356
      | | | | | | | : | | : | | : | : | | | | | | | : |
Db     404 YTGPACSQDVDECALGAN-----PCEHAGKCLNTLGSFECQLQGYTGPRCEIDV-- 453

Qy     357 EGLYGIKCDKRC---PCHLENTHSCHPMSG--CACKPGWSGLYC----- 396
      | | | : | : | | | | | | | | | | | | | | | | | | |
Db     454 -----NECISNPC--QNDATCLDQIGEFQCICMPGYEGVYCEINTDECASSPCLHN 502

Qy     397 -----NE---TCSPGFYGEACQ---QICS---CQNGADC--DSVTGKCTCAPGFKGID 438
      | | | | | | | | : | : | | | | | | | | | | | | : |
Db     503 GRCVDKINEFLCQCPKGFSGHLCQYDVDECASTPCKNGAKCLDGPNTYTCVCTEGYTGTGTH 562

Qy     439 CST-----PCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGW----- 477
      | | : | | | | : | : | : | : | : | : | : | : |
Db     563 CEVDIDECDPDPCCHIGL-----CK-DGVAT---FTCLCQPGYTGHHCETNINECH 608
```

Qy 478 -----HGVDCSIR-----CPSGTWGFGCNLTCQ-----CLNGGACNTLDG--TCTCAP 518
 || | | | | | | : | : | : | | |
 Db 609 SQPCRHHGGTCQDRDNYLCLCLKGTTGPNCEINLDDCASNPCDSGTCLDKIDGYECACEP 668
 Qy 519 GWRGEKCEL-----PCQDGTGYNCAERCDCHADGCHPTT----- 554
 |: | | : | : | | | | : | | |
 Db 669 GYTGSMCNVNIDECAGSPCHNGGTCEDEGIAGFTC--RC----PEGYHDPTCLSEVNECNS 722
 Qy 555 -----GHCRC-----CLPGWSGVHCD-----SVCAEG 575
 | || | |||| : || | ||
 Db 723 NPCIHGACRDGLNGYKCDCAPGWSGTNCDINNNECESNPCVNGGTCKDMTSGYVCTCREG 782
 Qy 576 RWGPNC-----SLPCY-----CKNGA 591
 |||| | || ||||
 Db 783 FSGPNCQTNINECASNPCLNQGTICDDVAGYKCNCPYPYTGATCEVVLAPCATSPCKNSG 842
 Qy 592 SCSPDDGI---CECAPGFRGTTTCQ-----RICSPGFYGH 622
 | : | | | : | : | : | : | : |
 Db 843 VCKESEDYESFSCVCPTGWQGTCEIDINECVKSPCRHGASCQNTNGSYRCLCQAGYTGR 902
 Qy 623 RCSQTCPCQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
 | | | | : | | | | | | | | | |
 Db 903 NCESDIDDC--RPNPCHNGGSCTDGVNAAFCDCLPGFQGAFCCEEDINECATNPCQNGANC 960
 Qy 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPGWIGSDC 699
 ||: | | : | | | | | : | : | | : | | |
 Db 961 TDCVDSYTCTCPTGFENGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGFTGSYC 1018
 Qy 700 -----SQP-----CPPAHWGPNC---IHTCN---CHNGAFC--- 724
 |: | | | | : | : | : | | | |
 Db 1019 QYDVNECDSPCLHGGTCQDSYGTGKCTCPQGYTGLNQCQNLVRWCDSAPCKNGGKCWQTN 1078
 Qy 725 SAYDGECKCTPGWTGLYC---TQRCPLGFY--GKDCALICQ----- 760
 : | | : | | | | : | : | | : | |
 Db 1079 TQY--HCECRSGWTGFNCQVLSVSCQVAAQKRGIDVTLLCQHGGLCVDEEDKHYCHCQAG 1136
 Qy 761 -----CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 |||| | | : | | | : | : | :
 Db 1137 YTGSYCEDEVDECSNPNCQNGATCTDYLGGSCKCVAGYHGSNCSEEINECLSQPCQNGG 1196
 Qy 789 -----CPSGTGYGCRQICD-----CLNNSTC-DHITG-TCYCS 820
 || || | | | | | | | | : | | |
 Db 1197 TCIDLNTYKCSGPRGTQGVHCEINVDCHPPLDPASRSPKCFNNGTCVDQVGGYTCTCP 1256
 Qy 821 PGWKGARCDQAGVIVGNLN 840
 ||: | ||: | : |
 Db 1257 PGFVGECRCE-----GDVN 1269

RESULT 8

A49128

cell-fate determining gene Notch2 protein - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002

C;Accession: A49128

R;Weinmaster, G.; Roberts, V.J.; Lemke, G.

Development 116, 931-941, 1992

A;Title: Notch2: a second mammalian Notch gene.
 A;Reference number: A49128; MUID:93202015; PMID:1295745
 A;Accession: A49128
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: mRNA
 A;Residues: 1-2471 <WEI>
 A;Experimental source: Schwann cell
 A;Note: sequence extracted from NCBI backbone (NCBIP:127811)
 C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F;264-295/Domain: EGF homology <EGX1>
 F;799-830/Domain: EGF homology <EGF1>
 F;877-908/Domain: EGF homology <EGX2>
 F;1029-1060/Domain: EGF homology <EGF>
 F;1067-1098/Domain: EGF homology <EGX3>
 F;1153-1184/Domain: EGF homology <EGF3>
 F;1191-1222/Domain: EGF homology <EGX4>
 F;1876-1908/Domain: ankyrin repeat homology <AN1>
 F;1909-1941/Domain: ankyrin repeat homology <AN2>
 F;1943-1975/Domain: ankyrin repeat homology <AN3>
 F;1976-2008/Domain: ankyrin repeat homology <AN4>
 F;2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 14.8%; Score 998; DB 2; Length 2471;
 Best Local Similarity 24.4%; Pred. No. 2.6e-45;
 Matches 321; Conservative 79; Mismatches 322; Indels 596; Gaps 70;

```

Qy      17  CHWIGTASPLNLEDPNVCSHWESYSVTVQESY-----PHFPDQIYYTSCTDIL 64
      || :          |::| | | :          || :      ::|| :
Db      121 CHMLS-----WDTYECTCQVGFTGKQCQWTDVCLSHPCEN--GSTCSSVA 163

Qy      65  NWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTC 124
      | | |          :| | :|| | :      : ||| ||
Db      164 NQFSC-----RCPAGI--TGQKCDADINECDIPGRCQHGGTC 198

Qy      125 QCEPGWGGTNCSSACDGDHWGPHCTS-----RCQCKNGALC---NPITGACHCAAGFRG 175
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      199 LNLP--GSYRCQ--CPQRFTHGQCDSPYVPCAPSPCVNGGTCRQTGDFTSECHCLPGFEG 254

Qy      176 WRCEDRCEQGTYGNDCHQRCQCQNGATC-DHV-TGECRCPPGYTGAFD-EDLCPPGKHGP 232
      || :      || :||| | | | | | |||| :|| | |||
Db      255 SNCERNID-----DCPNH-KCQNGGVCDVGVNTYNCRCPPQWTGQFCTEDV-----D 300

Qy      233 QC-EQRCPCQNGGVCHHVTG--ECSCPSGWMGTVCQP----- 267
      :| | | |||| | : | | | :|| | | :
Db      301 ECLLPNACQNGGTCTNRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASF 360

Qy      268 ---CPEGRFGKNC--SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCQ---DECP 312
      ||||: | | | | | | | | | | | | | | | | | |
Db      361 SCLCPEGKAGLLCHLDDACISNPCHKGALCDTNPLNGQYICTCPQAYKGADCTEDVDECA 420

Qy      313 VGTYGVLCAETCQCVNGGKCYHVSGA--CLCEAGFAGERCE----- 351
      :      | : | : ||| : || | | | :|| |||
Db      421 M-----ANSNPCEHAGKCVNTDGAHFHCECLKGYAGPRCEMDINECHSDPCQNDATCLD 473

Qy      352 -----ARLCPEGLYGIKCDKR-----CP-----CHLE- 373
      || | | : | :          ||          | :
Db      474 KIGGFTCLCMPGFGKVHCELEVNECQSNPCVNNQCVDKVNRFQCLCPPGFTGPVCQIDI 533
  
```

Qy	374	-----NTHSC--HPMSGECACKPGWSGLYCNET-----	399
Db	534	DDCSSTPCLNGAKCIDHPNGYECQCATGFTGTLCDENIDNCDPDPCHHGQCQDGDIDSYTC	593
Qy	400	-CSPGFYGEAC--QQICSC-----QNGADCDSVTG-KCTCAPGFGKIDC-----STP	442
		: : : : :	
Db	594	ICNPGYMGAI CSDQIDECYSSPCLNDGRCIDL VNGYQCNCQPGTSGLNCEINFDDCASNP	653
Qy	443	CPLGTY--GIN-----CS-----SRCG-----CKNDAVC-----	464
		:	
Db	654	CLHGACVDGINRYSCVCSPGFTGQRCNIDIDECASNPCRKDATCINDVNGFRCMCPEGPH	713
Qy	465	-----SP-VDGSCT-----CKAGWHGVDCSI-----	484
		: : : :	
Db	714	HPSCYSQVNECLSSPCIHGNCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGGTCN	773
Qy	485	-----RCPSGTWGFGCNLTQ-----CLNGGAC-----	507
		: :	
Db	774	NLVNGYRCTCKKGFGKYNCQVNI DECASNPCLNQGTCLDDVSGYTCHCMLPYTGKNCQTV	833
Qy	508	-----NTLDGTCTCAPGWRGEKCELPQDGTGYLNCAERCDCSHAD	548
		: : : : :	
Db	834	LAPCSPNPCENAAVCKEAPNFESFTCLCAPGWQGQRCTVDVDE-----CVSK-PCMNG	886
Qy	549	GCHPTTGH--CRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASC--SPDDGICECAP	604
		: : : :	
Db	887	ICHNTQGSYMECEPPGFGMDCEE-----DINDCLANPCQNGGSCVDKVNFTFSCLCPL	939
Qy	605	GFRGTTQQR-----ICSPGFYGHRCSTCPQCVHSS-----	635
		: :	
Db	940	GFVGDKCQTDMNECLSEPCKNGGTCSDYVNSYTC TCPAGFHGVHCENNIDECTESSCFNG	999
Qy	636	GPCHHITGL---CDCLPGFTGALC---NE-----VCPSGRFG	666
		: :	
Db	1000	GTC--VDGINSFSCLCVPVGFTGPFCLHDINECSSNPCLNSGTCVDGLGTYRCTCPLGYTG	1057
Qy	667	KNC---AGICT---CTNNGTC--NPIDRSCQCYPGWIGSDCSQ-----	701
		: : :	
Db	1058	KNCQTLVNLCSPPCKNKGTCAQEKARPRCLCPPGWDGAYCDVLNVSCKAAALQKGVPE	1117
Qy	702	-----PCPPAHWGPNC---IHTC---NCHNGAFCSAYDG--ECKCTP	735
		: : : : :	
Db	1118	HLCQHSGICINAGNTHHCQCPLGYTGSYCEEQLDECASNPCQHGATCSDFIGGYRCECVP	1177
Qy	736	GWTGLYCTQR-----CPLGFYG-----KDCALICQCQN	763
		: :	
Db	1178	GYQGVNCEYEVEDECQNQPCQNGGTCIDL VNHFKCSCPPGTRGLLCEENIDDCAGAPHCLN	1237
Qy	764	GADC-DHISG-----QCTCRTGFM	781
		:	
Db	1238	GGQCVDRIGGYSCRCLPGFAGERCEGDINECLSNPCSSEGLDCIQLKNNYQCVCRSAFT	1297
Qy	782	GRHCE---QKCPSTYGYGCRQICDLNNSTCDHITGT-----CYCSPGWKGARCDQA	831
		: : :	
Db	1298	GRHCETFLDVCPQK-----PCLNGGTCAVASNVDPGFICRCPGFGSARCQSS	1345

RESULT 9

S42612

transmembrane protein precursor - zebra fish

C;Species: Brachydanio rerio (zebra fish)

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002

C;Accession: S42612

R;Bierkamp, C.; Campos-Ortega, J.A.

Mech. Dev. 43, 87-100, 1993

A;Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern of transcription during early embryogenesis.

A;Reference number: S42612; MUID:94128602; PMID:8297791

A;Accession: S42612

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-2437 <BIE>

A;Cross-references: EMBL:X69088; NID:g433866; PIDN:CAA48831.1; PID:g433867

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;755-786/Domain: EGF homology <EGF1>

F;1023-1054/Domain: EGF homology <EGF>

F;1185-1216/Domain: EGF homology <EGF2>

F;1915-1947/Domain: ankyrin repeat homology <AN1>

F;1948-1980/Domain: ankyrin repeat homology <AN2>

F;1982-2014/Domain: ankyrin repeat homology <AN3>

F;2015-2047/Domain: ankyrin repeat homology <AN4>

F;2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 14.6%; Score 987; DB 2; Length 2437;

Best Local Similarity 24.8%; Pred. No. 9.7e-45;

Matches 310; Conservative 81; Mismatches 320; Indels 538; Gaps 70;

```

Qy      91 KSQCCPGFYESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGGTNCSSACD 140
      | | || | :|: |:: | : ||:|:| | | |
Db      85 KCDCVLGF--SDRLCLTPVNHACMNSPCRNGGTCSLLTLDFTTCRCQPGWSGKTCQLA-- 140

Qy     141 GDHWGPHCTSRCQCKNGALCNPITG--ACHCAAGFRGWRCEDRCEQGTYGNDCH-QRCQC 197
      | | | || |: | | | | | | | | | |
Db     141 -----DPCASN-PCANGGQCSAFESHYICTCPPNFHGQTCRQDV-----NECAVSPSPC 188

Qy     198 QNGATCDHVTGE--CRCPPGYTGAFCEDL-----CPPGK 229
      :|| || : | |||| || | : | | |
Db     189 RGGGTCINEVGSYLCRCPPEYTGPHCQRLYQPCLPSPCRSGGTCVQTSDDTHTCSCLPGF 248

Qy     230 HGPQCE-----QRC-----PCQNG 243
      | || | | | | | | | | | |
Db     249 TGQTCEHNVDDCTQHACENGGPICIDGINTYNCHCDKHWTGQYCTEDVDECESPNACQNG 308

Qy     244 GVCHHVTG--ECSCPSGWMGTVCQG-----PCPEGRFGKN 276
      | ||: | | | :|| | | : | | | |
Db     309 GTCHNTIGGFHCVCVNGWTGDDCSENIDDCASAAACSHGATCHDRVASFFCECPHGRTGLL 368

Qy     277 C--SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETC 324
      | | | | | | | | | | | | | | | |
Db     369 CHLDDACISNPCQKGSNCDTNPVSGKAICTCPPGYTGSAQNQDIDECSLGN----- 420

Qy     325 QCVNGGKCYHVSQA--CLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMS 382
      | :||:| : |: | | :| | || | : | | | |

```

Db 421 PCEHGGRCCLNTKGSFQCKCLQGYEGPRCEMDV-----NEC-KSNPC--QNDATCLDQI 470
 Qy 383 G--ECACKPGWSGLYCNET-----CSPGFYGEACQ---QIC 413
 | | | | : | : | | | | | | | | | |
 Db 471 GGFHCICMPGYEGVFCQINSDDCASQPCLNGKCIDKINSFHCECPKGFSGSLCQVDVDEC 530
 Qy 414 S---CQNGADCDSVTGK--CTCAPGFKGIDC-----STPCPL-----GTYGINCSSR 455
 : | : | | | | | | | | | | : | | | | |
 Db 531 ASTPCKNGAKCTDGPKNYTCECTPGFSGIHCELDINECASSPCHYGVCRDGVASFTCD CR 590
 Qy 456 CG-----CKNDAVCSPVDGS--CTCKAGWHGVDCSIR----- 485
 | | | | : | : | | | | | | : | | | |
 Db 591 PGYTGRLCETNINECLSQPCRNGGTCQDRENAYICTCPKGTTGVNCEINIDDCRKRKPCDY 650
 Qy 486 -----CPSGTWGFGCNLTQ-----CLNGGACNTLDG---TCTCAPGWRG 522
 | | | | : | | | | | : | | | | : |
 Db 651 GKCIDKINGYECVCEPGYSGSMCNINIDDCALNPCHNGGTC--IDGVNSFTCLCPDGFRD 708
 Qy 523 EKC-----ELPCQDGTYGLNC---AERC---DCSHADGCHP 552
 | | | | : | | | | | | | : | | |
 Db 709 ATCLSQHNECSSNPCIHGSCLDQINSYRCVCEAGWMGRNCDININECLSNPCVNGGTCKD 768
 Qy 553 -TTGH-CRCLPGWSGVHCD SVCAEGRWGP-----NCSL----- 583
 | : | : | | | : | : | | | | | | | | | |
 Db 769 MTSGYLCTCRAGFSGPNCQMNINECASNPCLNQGSCIDDVAGFKCNCMLPYTGEVCENVL 828
 Qy 584 -PCY---CKNGASCSPDD-----GICE----- 601
 | | | | | | : | | | | | : | | | |
 Db 829 APCSPRPCKNGGVCRESEDFQSFSCNCPAGWQGQTCEVDINECVRN PCTNGGVCCENLRGG 888
 Qy 602 ----CAPGFRGTTCQR-----ICSPGFYGHRCSTCPQCV 632
 | | | | | : | | | | | : | | | | : | : | | |
 Db 889 FQCR CNPGFTGALCENDIDDCEPNPCSNNGGVCQDRVNGFVCVCLAGFRGERCAEDIDECV 948
 Qy 633 HSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNC---AGICT---CTNNGTCNPIDR 686
 | | | : | : | : | : | | | | | | | | | : |
 Db 949 --SAPCRNNGNCTDCVNSYT--CS--CPAGFSGINCEINTPDCTESSCFNGGTC--VDG 999
 Qy 687 ----SCQCYPGWIGSDC-----SQP-----CPPAHWGPNC---IH 714
 | | | | : | : | | | : | | | | : | | | : |
 Db 1000 ISSFSCVCLPGFTGNYCQHVDVNECD SRPCQNGGSCQDGYGTYKCTCPHGYTGLNCQSLVR 1059
 Qy 715 TCN---CHNGAFC--SAYDGECKCTPGWTGLYCTQ----- 744
 | : | | | | | : | | | | : | | | | |
 Db 1060 WCDSSPCKNGGSCWQOGASFTCQCASGWTGIYCDVPSVSCEVAARQQGVSVAVLCRHAGQ 1119
 Qy 745 -----RCPLGFYKDC--ALICQ---CQNGADC-DHISG-QTCRTGFMGRHCE 786
 | | | : | | | | | | | | | : | | | : | : |
 Db 1120 CVDAGNTHLCRCQAGYTGSYCQEQVDECQPNPCQNGATCTDYLGGYSCECVPGYHGMNCS 1179
 Qy 787 QK-----CPSGTYGYGCRQICD-----CLNN 807
 : : | | | | | | | | | | | | | | | |
 Db 1180 KEINECLSQPCQNGGTCIDL VNTYKCS CPRGTQGVHCEIDIDDCSPSVDPLTGEPRCFNG 1239
 Qy 808 STC-DHITG-TCYCSPGWKGARCDQAGVII VGNLNSLSRTSTALPADSY 854
 | | : | | | | : | | : | : | : | : | : |
 Db 1240 GRCVDRVGGYGCVC PAGFVGERCE-----GDVNE-CLSDPCDPSGSY 1280

RESULT 10

A24420

notch protein - fruit fly (*Drosophila melanogaster*)

N;Alternate names: neurogenic repetitive locus protein

C;Species: *Drosophila melanogaster*

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C;Accession: A24420; A24768; S09358; A05267

R;Kidd, S.; Kelley, M.R.; Young, M.W.

Mol. Cell. Biol. 6, 3094-3108, 1986

A;Reference number: A24420; MUID:87064624; PMID:3097517

A;Accession: A24420

A;Molecule type: DNA

A;Residues: 1-2703 <KID>

A;Cross-references: GB:K03508; NID:g157991; PIDN:AAA28725.1; PID:g157993

R;Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.

Cell 43, 567-581, 1985

A;Reference number: A24768; MUID:86079539; PMID:3935325

A;Accession: A24768

A;Molecule type: mRNA

A;Residues: 1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',268-
872,'R',874-958,'R',960-1970,'FH',1973-2256,'G',2258-2264,'V',2266-
2406,'R',2408-2444,'L',2446-2703 <WHA1>

A;Note: the authors translated the codon ATC for residue 49 as Thr, ATT for
residue 2044 as Arg, GTA for residue 2265 as Ala, CGC for residue 2407 as His,
and CTT for residue 2445 as Arg

R;Tautz, D.

Nucleic Acids Res. 17, 6463-6471, 1989

A;Title: Hypervariability of simple sequences as a general source for
polymorphic DNA markers.

A;Reference number: S09358; MUID:89385974; PMID:2780284

A;Accession: S09358

A;Molecule type: DNA

A;Residues: 2505-2551,'QQQQ',2552-2576,'E',2578-2604 <TAU>

R;Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.

Cell 40, 55-62, 1985

A;Title: opa: a novel family of transcribed repeats shared by the Notch locus
and other developmentally regulated loci in *D. melanogaster*.

A;Reference number: A05267; MUID:85099329; PMID:2981631

A;Accession: A05267

A;Molecule type: DNA

A;Residues: 2504-2576,'E',2578-2611 <WHA2>

C;Genetics:

A;Gene: notch; opa

A;Cross-references: FlyBase:FBgn0004647

A;Map position: 8.96-9.36

A;Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

C;Keywords: differentiation; tandem repeat; transmembrane protein

F;27-43/Domain: transmembrane #status predicted <TMM1>

F;297-328/Domain: EGF homology <EGX1>

F;530-561/Domain: EGF homology <EGF1>

F;568-599/Domain: EGF homology <EGF>

F;988-1019/Domain: EGF homology <EGX2>

F;1064-1095/Domain: EGF homology <EGF3>

F;1187-1218/Domain: EGF homology <EGX3>

F;1746-1762/Domain: transmembrane #status predicted <TMM2>
 F;1950-1982/Domain: ankyrin repeat homology <AN1>
 F;1983-2015/Domain: ankyrin repeat homology <AN2>
 F;1988-2004/Domain: transmembrane #status predicted <TMM3>
 F;2017-2049/Domain: ankyrin repeat homology <AN3>
 F;2050-2082/Domain: ankyrin repeat homology <AN4>
 F;2083-2115/Domain: ankyrin repeat homology <AN5>
 F;2538-2568/Region: glutamine-rich
 F;2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 14.5%; Score 978.5; DB 1; Length 2703;
 Best Local Similarity 26.8%; Pred. No. 3e-44;
 Matches 290; Conservative 102; Mismatches 297; Indels 395; Gaps 70;

```

Qy      7 SCL----SFICLLLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
      |||  :| |: : : ||  ::::  | |  :  :|
Db    502 SCLDDPGTFRCVCMPTGTGTQCEIDIDECQSNPC-----LNDGTC 541

Qy      61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMC---VPHCADKCVHGR 117
      | :| |||:  | || :| | : | : |
Db    542 HDKINGFKCS-----CALGF--TGARCQINIDDCQSQPCRNR 576

Qy     118 -----CIAPNTCQCEPGWGGTNCS---SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH 168
      || :|:| ||: ||:| : || :  | | : : |
Db    577 GICHDSIAGYSCECPPGYTGTSCEININDCDSN-----PCHRGKCIDDVNSFKCL 626

Qy     169 CAAGFRGW-----RCEDR-----CEQGTYG-----NDCHQRCQ 196
      | |: |:  |:||  |: || |  |: ||
Db    627 CDPGYTGyticQKQINECESNPCQFDGHCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P 685

Qy     197 CQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVC-HHVTG-E 252
      | ||||| | : :|:| ||:| ||  | : :|  || | |||  | | :
Db    686 CNNGATCIDGINSYKQCVPFGFTGQHCE-----KNVDECIS-SPCANNGVCIDQVNGYK 738

Qy     253 CSCPSGWMGTVC-----GQP-----CPEGRFGKNCS---QECQ-- 282
      | || |:  |  |  || | || |  ||
Db    739 CECPRGFYDAHCLSDVDECASNPCVNEGRCEDEGINEFICHCPPGYTGKRCELDIDECSSN 798

Qy     283 -CHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNNGGKCY-HV 335
      | :||| |  | | |||||:|: | :|  | | ||| |  |
Db    799 PCQHGGTCYDKLNAFSCQCMPTGTGQKCEITNIDDC-----VTNPGCNGGTCTIDKV 848

Qy     336 SG-ACLCEAGFAGERCEARLCPEGLYGIKC-DKRCPCHLNTHSCHPMSG---ECACKP 389
      :| |:|:  | | ||::|  | || :|  | | |  | ||
Db    849 NGYKCVCKVPFTGRDCESKMDP-----CASNRC---KNEAKCTPSSNFLDFSCCTCKL 897

Qy     390 GWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTG--KCTCAPGFKGIDC----- 439
      |::| ||:|  : |  |:||| | :| |  :| |  |:| ||
Db    898 GYTGRYCDEDI-----DECSSLSSPCRNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCA 951

Qy     440 STPCP-----LGTYGINC-----SSRC-----GCKNDAVCSVPDGS--CTCK 474
      | ||  :| |  |  |  |:| || |  | |||
Db    952 SFPCQNGGTCLDGIGDYSLCVDGFDGKHCETDINECLSQPCQNGATCSQYVNSYTCTCP 1011

Qy     475 AGWHGVDCSIR-----CPSGTWGFGCN-----LTCQ-----CLN 503
      | : |:|  |  | |: | :  ||  |||
Db   1012 LGFSGINCQTNDEDCTESSCLNGGSCIDGINGYNCSLAGYSGANCQYKLNKCDSNPCLN 1071

```

Qy	504	GGACNTLDG--TCTCAPGWRGEKC-----ELPCQDGTGYGLNCAERCDSCSHADGCHPT	553
Db	1072	: : : : : ::	
		GATCHEQNNEYTCHCPSGFTGKQCSEYVDWCGQSPCENG-----ATCSQMK--HQF	1120
Qy	554	TGHCRCPLPGWSGVHCD--SVCAEGRWGPNCSLPCYCKNGASCSP--DDGICECAPGFRG	608
Db	1121	: : : : : : : : :	
		S--CKCSAGWTGKLCDVQTI SCQDAADRKGLSLRQLCNNG-TCKDYGNSHVCYCSQGYAG	1177
Qy	609	TTCQR-----ICSPGFYGHRCSTCPQCV---HSSGPCH	639
Db	1178	: : :	
		SYCQKEIDECQSQPCQNGGTCRDLIGAYECQCRQGFQGNCELNIDDCAPNPCQNGGTCH	1237
Qy	640	H--ITGLCDCLPGFTGALC---NEVCPSGRFGKNCAGICTCTNNGTCNPIDR---SCQC	690
Db	1238	: : :	
		DRVMMNFSCSCPFGMTGIIICEINKDDCKPG-----ACHNNGSC--IDRVGGFECVC	1285
Qy	691	YPGWIGSDC-----SQPCP-----PAHWGPNCIHTCN-----	717
Db	1286	:: : :	
		QPGFVGARCEGDINECLSNPCSNAGTLDVCVQLVNNYHCNCRPGHMGRHCEHKVDFCAQSP	1345
Qy	718	CHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHIS---GQC	774
Db	1346	: : : : :	
		CQNGGNCNIRQ-----SGHHCI--CNNGFYGKNCEL-----SGQDCDSNPCRVGNC	1389
Qy	775	TCRTGFMGRHCEQKCPSTYGYGCR---QICD---CLNNSTCDHITG--TCYCSPGWKG	825
Db	1390	: : :	
		VVADEGFGYRCE--CPRGTLGEHCEIDTLDECSNPPCAQGAACEDLLGDYECLCPSKWKG	1447
Qy	826	ARCD 829	
Db	1448		
		KRCD 1451	

RESULT 11

S45306

notch 3 protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 20-Feb-1995 #sequence revision 20-Feb-1995 #text change 02-Aug-2002

C;Accession: S45306

R;Lardelli, M.; Dahlstrand, J.; Lendahl, U.

Mech. Dev. 46, 123-136, 1994

A;Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-repeats and is expressed in proliferating neuroepithelium.

A;Reference number: S45306; MUID:95001556; PMID:7918097

A;Accession: S45306

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-2318 <LAR>

A;Cross-references: EMBL:X74760; NID:q483580; PIDN:CAA52776.1; PID:q483581

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

F;163-195/Domain: EGF homology <EGF1>

F;474-505/Domain: EGF homology <EGF>

F;854-885/Domain: EGF homology <EGF2>

F;1839-1871/Domain: ankyrin repeat homology <AN1>

F;1872-1904/Domain: ankyrin repeat homology <AN2>

F;1906-1938/Domain: ankyrin repeat homology <AN3>

F;1939-1971/Domain: ankyrin repeat homology <AN4>
F;1972-2004/Domain: ankyrin repeat homology <AN5>

Query Match 14.5%; Score 977.5; DB 2; Length 2318;
Best Local Similarity 25.8%; Pred. No. 3e-44;
Matches 322; Conservative 77; Mismatches 317; Indels 532; Gaps 74;

```
Qy      9 LSFICLLLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNWFK 68
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      62 LEAACLCLPGWVG--ERCQLEDP--C-----HSGPCAGRGVCQSSVVAGTARFS 106

Qy      69 CTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCV---PHCADKCVHGR--CIAPN- 122
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     107 C-----RCLRGF--QGPDCSQPDPCVSRPCVHGAPCSVGPDPG 141

Qy     123 --TCQCEPGWGGTNCSSACDGDHWGPHC-----TSRCQ----- 153
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     142 RFACACPPGYQGQSCQSDIDECSRGTTCRHGGTCLNTPGSFRCQCPLGYTGLLCENPVVP 201

Qy     154 -----CKNGALC---NPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRQCQNGATC-D 204
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     202 CAPSPCRNGGTCRQSSDVITYDCACLPGFEGQNCENVND-----DCPGH-RCLNGGTCVD 254

Qy     205 HV-TGECRCPPGYTGAF-C-EDLCPPGKHGPQCE-QRCPCQNGGVCHHVTG--ECSCPSGW 259
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     255 GVNTYNCQCPPEWTGQFCTEDV-----DECQLQPNACHNGGTFCFNLLGGHSCVCVNGW 307

Qy     260 MGTVCQG-----PCPEGRFGKNC--SQEC---QCHNGGTC 289
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     308 TGESCSQNIDDCATAVCFHGATCHDRVASFYCACPMGKTGLLCHLDDACVSNPCHEDAIC 367

Qy     290 DA--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLC AETCQCVNNGGKCYHVS GA--CL 340
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     368 DTNPVSGRAICTCPPPGFTGGACDQDVDECSIG-----ANPCEHL--GRCVNTQGSFLCQ 419

Qy     341 CEAGFAGERCEARL--CPEGLYGIKCDKRCPCCHLENTHSCHPMSG--CACKPGWSGLYC 396
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     420 CGRGYTGPRCETDVNECLSG-----PC--RNQATCLDRIGQFTCICMAGFTGTYC 467

Qy     397 NE-----TCSPGFYGEACQ---QICS---CQNGADC-DSV 424
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     468 EVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSMCQLDVDECASTPCRNGAKCVDQP 527

Qy     425 TG-KCTCAPGFKGI-----DCS-TPCPLGTYGINCSSRCGCKNDAVCSPVDG---SC 471
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     528 DGYECRCAEGFEGTLCERNVDDCSPDPCHHG-----RC-----VDGIASFSC 569

Qy     472 TCKAGWHGVDCS-----IRCPSGTWGFGCNLT-CQLNG- 504
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     570 ACAPGYTGIRCESQVDECRSQPCRYGGKCLDLVDKYLRCPPGTTGVNCEVNIDDCASNP 629

Qy     505 ---GACNTLDG---TCTCAPGWRGEKCEL-----PCQDGTGYGLNC----- 538
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     630 CTFGVCR--DGINRYDCVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFHCLCPPG 687

Qy     539 -----AERCDCSHADGCHPTTG--HCRCLPGWSGVHCDSDVCAEGRWGPNC SLP 584
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
```


Db 688 SLPLCLPANHPCAHKPCSHG-VCHDAPGGFRCVCEPGWSGPRCSQSLA-----PDACES 741
 Qy 585 CYCKNGASCSPDDGI---CECAPGFRGTTCQRI-----CS 616
 | : | : | : | | | | | : | : : |
 Db 742 QPCQAGGTCT-SDGIGFRCTCAPGFQGHQCEVLSPTPSLCEHGGHCESDPDRLTVCSCP 800
 Qy 617 PGFYGHRCSQTCPQCVHSS-----GPCHHITG--LCDCLPGFTGALCNE----- 658
 | : | | | : | : | | : | | | : | : :
 Db 801 PGWQGPRCQQDVDECA GASPCGPHGTCTNLPGNFRCICHRGYTGPFCDQDIDDCDENPCL 860
 Qy 659 -----VCPSGRFGKNCA-----GICT----- 674
 | | | | | | |
 Db 861 HGGSCQDGVGSFSCSCLDGFAGPRCARDVDECLSSPCGPGTCTDHVASFTCACPPGYGGF 920
 Qy 675 -----CTNNGTCNPIDR---SCQCYPGWIGSDC-----SQP-----C 703
 | | | | : | | | | : | : | : | : |
 Db 921 HCEIDLPCDSPSSCFNGGTC--VDGVSSFSLCRPGYTGTGHCQYEADPCFSRPCLHGGIC 978
 Qy 704 PPAHWGPNCIHTCN-----CHNGAFCSAYDGECKCTPGWTG----- 739
 | | | | | | | | | | | : |
 Db 979 NPTHPGFEC--TCREGFTGSQCQNPVDWCSQAPCQNGGRCVQTGAYCICPPGWSGRLCDI 1036
 Qy 740 --LYCTQR-----CPLGFYKDC-----ALICQCQN 763
 | | : | | | | | :
 Db 1037 QSLPCTEAAAQMGVRLEQLCQEGGKCIDKGRSHYCVCEGRTGSHCEHEVDPCTAQPCQH 1096
 Qy 764 GADCDHISG--QCTCRTGFMGRHCEQ-----KCPSGTYGY 796
 | | | | | : | | | | |
 Db 1097 GGTCTRGYMGGYVCECPAGYAGDSCEDNIDECASQPCQNGGSCIDLVARVLCSCPPGTLGV 1156
 Qy 797 GC---RQICD-----CLNNSTCDHITG--TCYCSPGWKGARCD 829
 | | | | : | | | : | : | :
 Db 1157 LCEINEDDCDLGPSLD SGVQLHNGTCVDLVGGFRCNCPPGYTGLHCE 1204

RESULT 12

S78549

notch3 protein - human

C;Species: Homo sapiens (man)

C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 08-Sep-2002

C;Accession: S78549; S71825

R;Joutel, A.; Tournier-Lasserre, E.

submitted to the EMBL Data Library, April 1997

A;Reference number: S78549

A;Accession: S78549

A;Molecule type: mRNA

A;Residues: 1-2321 <JOU1>

A;Cross-references: EMBL:U97669; NID:g2668591; PIDN:AAB91371.1; PID:g2668592

R;Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.;

Alamowitch, S.; Domenga, V.; Cecillion, M.; Marechal, E.; Maciazek, J.;

Vayssiere, C.; Cruaud, C.; Cabanis, E.A.; Ruchoux, M.M.; Weissenbach, J.; Bach,

J.F.; Bousser, M.G.; Tournier-Lasserre, E.

Nature 383, 707-710, 1996

A;Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia.

A;Reference number: S71825; MUID:97032728; PMID:8878478

A;Accession: S71825

A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 67-113;138-194;268-333,'G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2>
A;Cross-references: EMBL:U97669
C;Genetics:
A;Gene: notch3
A;Map position: 19p13.1
C;Function:
A;Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and dementia
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
C;Keywords: tandem repeat; transmembrane protein
F;123-155/Domain: EGF homology <EGX1>
F;162-194/Domain: EGF homology <EGF1>
F;240-271/Domain: EGF homology <EGX2>
F;318-349/Domain: EGF homology <EGF>
F;473-504/Domain: EGF homology <EGX3>
F;853-884/Domain: EGF homology <EGF3>
F;928-959/Domain: EGF homology <EGX4>
F;1838-1870/Domain: ankyrin repeat homology <AN1>
F;1871-1903/Domain: ankyrin repeat homology <AN2>
F;1905-1937/Domain: ankyrin repeat homology <AN3>
F;1938-1970/Domain: ankyrin repeat homology <AN4>
F;1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 14.4%; Score 974; DB 2; Length 2321;
Best Local Similarity 25.0%; Pred. No. 4.5e-44;
Matches 304; Conservative 93; Mismatches 282; Indels 537; Gaps 70;

```

Qy      59 SCTDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKC----- 113
      :| | :| :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      250 TCVDGVNTYNC-----QCPPEW--TGQFCTED-VDECQLQPN 283

Qy      114 -VH--GRC---IAPNTCQCEPGWGGTNCSSACDG-----DHWGPHCTSR-----CQC--- 154
      | | | : :| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      284 ACHNGGTCFNTLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCHDRVASFYCACPMG 343

Qy      155 KNGALC-----NPITG--ACHCAAGFRGWRCE---DRCE----- 183
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      344 KTGLLCHLDDACVSNPCHEDAICDTNPVNGRAICTCPPGFTGGACDQDVDECSIGANPCE 403

Qy      184 -----QGTYGNDCHQ-----RCQ-----CQNGATCDHVTGE--CRCPPGYTG 218
      ||: : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      404 HLGRCVNTQGSFLCQCGRGYTGPRCETDVNECLSGPCRNQATCLDRIGQFTCICMAGFTG 463

Qy      219 AFCE---DLCPPGKHGPQCEQRCPCQNGGVC-HHVTG-ECSCPSGWMGTVC----- 264
      :|| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      464 TYCEVDIDEC-----QSSPCVNGGVCKDRVNGFSCTCPSGFSGSTCQLDVDECAS 513

Qy      265 -----GQP-----CPEGRFGKNCSQ---ECQ---CHNGGTCDA-ATGQCHCSPG 301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      514 TPCRNGAKCVDQPDGYECRCAEGFEGTLCDRNVDDCSPDPCHHGRCVDGIASFSCACAPG 573

Qy      302 YTGERCQDE-----CPVGTYGVLC-----AETCQ--- 325
      ||| ||: : | | | | | | | | | | | | | | | | | | | | | | | |
Db      574 YTGTRCESQVDECERSQPCRHHGGKCLDLVDKYLRCRPSGTTGVNCEVNIDDCASNPCTFGV 633

```

Qy 326 CVNGGKCYHVSGACLCEAGFAGERCEARL-----CPEGLYGIKCDKRCP--- 369
 | : | | : : | | : | : | | : | | : | |
 Db 634 CRDGINRYD----CVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFRC--LCPPGS 687

 Qy 370 ----CHLENTHSC-----HPMSGECACKPGWSGLYCNE----- 398
 | | : | | | | : | | | | : :
 Db 688 LPPLC-LPPSHPCAHEPCSHGICYDAPGGFRCVCEPGWSGPRCSQSLARDACESQPCRAG 746

 Qy 399 -----TCSPGFYGEACQOI--CS---CQNGADCDSVTGK---CTCAPGFKG-- 436
 | | | | | : : | : | : | | : | : | : |
 Db 747 GTCSSDGMGFHCTCPPGVQGRQCELLSPCTPNPCEHGGRCESAPGQLPVCSCPQGWQGPR 806

 Qy 437 -----IDCSTPCPLGTYGINCSSRCGCKNDAVCSFVDGSCCTCKAGWHGVDCSIRCPSGTW 491
 : | : | | : | | : : | | | | : | | |
 Db 807 CQQDVDECAGPAPCGPHGI-CTNLAG-----SFSCCTCHGGYTGPSCDQDIND--- 852

 Qy 492 GFGCNLTCQCLNGGACNTLDG----TCTCAPGWRGEKC-----ELPCQDGTYGLNCA 539
 | : | | | : | | | : | | : | | | | |
 Db 853 ---CDPN-PCLNGGSCQ--DGVGSFSCSCLPGFAGPRCARDVDECLSNPCGPGT---CT 902

 Qy 540 ERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGI 599
 : | : | | | : | | : | : | | | : | | :
 Db 903 D-----HVASFTCTCPPGYGGFHCEQDL-----PDCS-PSSCFENGCTCV--DGV 943

 Qy 600 ----CECAPGFRGTTCQR-----ICSPGFYGHRC----SQTCPQC----- 631
 | | | : | | | : | | | | | | | |
 Db 944 NSFSCLCRPGYTGAHCQHEADPCLSRPCLHGGVCSAAHPGFRCTCLESFTGPQCQTLVDW 1003

 Qy 632 ----VHSSGPCHHITGLCDCLPGFTGALCNE----- 658
 : | | | | : : | | :
 Db 1004 CSRQPCQNGGRCVQTGAYCLCPPGWSGRLCDIRSLPCREAAAQIGVRLEQLCQAGGQCVD 1063

 Qy 659 -----VCPSGRFGKNC---AGIC---TCTNNGTCNPI--DRSCQCYPGWIGSDC--- 699
 | | | | : | | : | | | : | | : | : |
 Db 1064 EDSSHYCVCPEGRTGSHCEQEVDPCLAQPCQHGGTCRGYMGGYMCECLPGYNGDNCEDDV 1123

 Qy 700 ----SQPC-----PPAHWGPNCIHTCNCH 719
 | | | | | | | | : | | |
 Db 1124 DECASQPCQHGGSCIDLVARYLCSCPPGTLGVLCEINEDDCGPGPPLDSGPRCLHNGTCV 1183

 Qy 720 N--GAFCSAYDGECKCTPGWTGLYCTQ-----RCPLG 749
 : | | | | : | | | | | | | | | |
 Db 1184 DLVGGF-----RCTCPPGYTGRLCEADINECRSGACHAAHTRDCLQDPGGGFRCLCHAG 1237

 Qy 750 FYGKDCALI---CQ---CQNGADCDDHISG-----QCTCRTGFMGRHCEQ----- 787
 | | | : | : | : | | | | | | | :
 Db 1238 FSGPRCQTVLSPCESQPCQHGGQCRPSPGPGGLTFTCHCAQPFWGPRCERVARSRELQ 1297

 Qy 788 -----KCPSTYGYGCRQI-----CDCLNNSTCDHIT--- 814
 | | | | | | | : : |
 Db 1298 CPVGVPCQQTTPRGPRCACPPGLSGPSCRSFPGSPPGASNASCAAAPCLHGGSCRPAFLAP 1357

 Qy 815 -GTCYCSPGWKGARCD 829
 | : | | | :
 Db 1358 FFRCACAQGWGTGPRCE 1373

RESULT 13

T31070

notch homolog - sea urchin (*Lytechinus variegatus*)

C;Species: *Lytechinus variegatus* (variegated urchin)

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000

C;Accession: T31070

R;Sherwood, D.R.; McClay, D.R.

Development 124, 3363-3374, 1997

A;Title: Identification and localization of a sea urchin Notch homologue: insights into vegetal plate regionalization and Notch receptor regulation.

A;Reference number: Z20966; MUID:97454256; PMID:9310331

A;Accession: T31070

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-2531 <SHE>

A;Cross-references: EMBL:AF000634; NID:g2570350; PID:g2570351; PIDN:AAB82088.1

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 14.3%; Score 964.5; DB 2; Length 2531;
Best Local Similarity 25.6%; Pred. No. 1.6e-43;
Matches 303; Conservative 92; Mismatches 312; Indels 477; Gaps 74;

```

Qy      6 NSCLSFICLLLCHWIGTASPLNLED--PNVCSHWESYSVTVQESYPHPFDQIYYTSCTDI 63
      | :  :| | :  | |      :| :| | | |      :  | |
Db      336 NTYGNFSCICVRGWEGQTCEINKDDCTPNPCQ-----FECECEDR 375

Qy      64 LNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRC----- 118
      :  ||||      | ||      :| :| |  | | :|  |
Db      376 VASEFKCT-----CPPG--RTGLLC--HLEDACMSNPCHHTAQ 408

Qy      119 ----IAPNT--CQCEPGWGGTNCSSACDGDHWGPHCTSRCQ--CKNGALCNPITG--ACH 168
      :  :  | |  | :  | || |  |      | :      | :| |  |  :|
Db      409 CSTSVVDGSFICDCATGYQGFNCSEDID-----ECSLSMDSICQSGGTCQNFDDGGWSCL 462

Qy      169 CAAGFRGWRC---DRCE-----QGTYG-----NDCHQRCQ 196
      | :|| | || |  | | :      :| |      | :|
Db      463 CSSGFTGSRCETDIDECDDPCYNGGTCLNKRGGYACICLTGFTGTLCETDINECSSN-P 521

Qy      197 CQNGATCDHVTG--ECRCPPGYTGAFCE-----DLCPPGK 229
      | ||| :|  :|| | | | || | | :      | | |
Db      522 CLNGASCFDITGRFECACLAGYTGTTCQVNIDDCQSSPCENGGTCIDGVNQFTCLCETGY 581

Qy      230 HGPQCE---QRC---PCQNGGVCHHVTG--ECSCPSGWMGTVC-----GQP----- 267
      | :||      |  || |||| | :|  :| :|| |  |  |
Db      582 EGHRCEMDSDECASRPCMNGGVCEDLIGFYQCNCVPVGTSGDNCEYNHYDCSSNPCVNDGT 641

Qy      268 -----CPEGRFGKNCSQ---ECQ---CHNGGTC-DAATG----- 294
      | ||  | || :  :| :  |||| | |  |
Db      642 CVDGINEYTCMCHEGYRGLNCEEDIDDCESRPCHNGGTCVDEVNGYHCLCPIGYHDPFCM 701

Qy      295 -----QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCV 327
      | |  |||| |  ||      :  |
Db      702 SNINECSSNPCVNGGSDHGVNEYSCECMAGYTGTRCTDDFDEC-----SSNPCQ 751

Qy      328 NGGKC--YHVSGACLCEAGFAGERCEARL-----CPEGLYG 361
      :|| |  |  | :|| :  | |  :      || :  |

```

Db	752	HGGTC	CDNRHAFYNCTCQAGYTGLNCEVNIDDCVDEPCLNGGICIDEVNSFQCVC	PQT	FVG	811
Qy	362	IKCD-KRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQ	-----	ICS		414
		:	:	:	:	:
Db	812	LLCETERSPC---	EDNQCQ-NGATCVYSEDIYAGYSCR--	CTSGFQGNFCDDDRNECLFSP		865
Qy	415	CQNGADCD	SVTG--KCTCAPGFKGIDC-----	STP-----	CPLGT	447
		:	:	:	:	:
Db	866	CRNGGSCTNLEGSFECSCLPGYDGPICEINIDECASGPCTNGGICTDLID	DDYFCSCQ	RGF		925
Qy	448	YGINC---	SSRC---GCKNDAVCSP-VDG-SCTCKAGWHGVDCSIRCPSGTWGFGCN	LTC		499
		:	:	:	:	:
Db	926	TGKNCQNDTDECLSSPCRN	GATCHEYVDSYTC	SCLVGFSGMHCEINDQDCT-----	TS	978
Qy	500	QCLNGGACNTLDG----	TCTCAPGWRGEKCEL-----	PCQDGT	YGLNCAER-----	541
		:	:	:	:	:
Db	979	SCLYGGTC--	IDGVNSYTC	CECVTGYTG	SNCQIEINECDSDPCENGA---	TCQDRFGSYSC 1033
Qy	542	-CD-----	CSH-ADGCHP-----	TTGH---CRCLPGWSGVHCD-----		569
		:	:	:	:	:
Db	1034	HCDVGFTGLNCEHVVQWCSPQNNPCYNGATCVAMGHLYECHCASNWI	GKLCDVPKVS	CDI		1093
Qy	570	-----	SVCAEGRWGPNC-----	SLPCYCKN---GASCSPDDGICECAPGFRGTT	C	611
		:	:	:	:	:
Db	1094	AASDKNVTRSELCLN---	GGTCIDATSSHSCLCQDGYTGSYCEVNIDECASAPCHNGGTC			1150
Qy	612	QR-----	ICSPGFYGHRCSQTC	PQCVHSSGPCHHITGLC-DCLPGFTGALCNEVCPSG		663
		:	:	:	:	:
Db	1151	TDGVYSYTC	SCLPGFEGPRCQQNINEC--	ASSPCHN-GGQCHDMVNGYT--	CS--CPAG	1202
Qy	664	RFGKNCA---	GIC---TCTNNGTCNPIDR----	SCQCYPGWIGSDC-----	SQPCPP	705
		:	:	:	:	:
Db	1203	TQGTDCSINLDDCYEGACYHGGVC--	IDQVGTYTCD	CPLG	FGVQHCEGDVNECLSNPCDP	1260
Qy	706	AHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLG	FGYKDCALICQ---	CQ		762
		:	:	:	:	:
Db	1261	V-GSQDCVQLINNY-----	QCVC	KPGYTGDCEQEIP-----	NCQNDPCQ	1299
Qy	763	NGADC--	DHISGQCTC	RTGFMGRHCEQK-----	CPSG	792
		:	:	:	:	:
Db	1300	NNGLC	LPSDEGYCDCLRGFTGVHCETKLTPCGTHPCQNEGTCMEYGD	DDFDDYT	CMCPSG	1359
Qy	793	TYGYGCR-----	QICDCLNNSTC--	DHITGTCYCS	PGWK	GARC 828
		:	:	:	:	:
Db	1360	VSGDNCEIDYNECASSPCINGGTCLDEY	GQYRCDC	PATWNGRNC		1403

RESULT 14

A40136

fibropellin Ia - sea urchin (*Strongylocentrotus purpuratus*)

N;Alternate names: epidermal growth factor homolog precursor

N;Contains: alternatively spliced fibropellin Ib (EGFI)

C;Species: Strongylocentrotus purpuratus (purple urchin)

C;Date: 13-May-1992 #sequence revision 17-Sep-1997 #text change 21-Jul-2000

C;Accession: A40136; B40136; C40136; A29316; A43131

R; Delgadillo-Reynoso, M.G.; Rollo, D.R.; Hursh, D.A.; Raff, R.A.

J. Mol. Evol. 29, 314-327, 1989

A;Title: Structural analysis of the uEGF gene in the sea urchin
 Strongylocentrotus purpuratus reveals more similarity to vertebrate than to
 invertebrate genes with EGF-like repeats.
 A;Reference number: A40136; MUID:90112459; PMID:2514273
 A;Accession: A40136
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-114
 A;Cross-references: GB:X17530; NID:g10225; PID:g667061
 A;Accession: B40136
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 181-251,329-370,'R',372-408,'RA',411-441 <DE2>
 A;Accession: C40136
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 'K',747-821,898-978 <DE3>
 R;Hursh, D.A.; Andrews, M.E.; Raff, R.A.
 Science 237, 1487-1490, 1987
 A;Title: A sea urchin gene encodes a polypeptide homologous to epidermal growth
 factor.
 A;Reference number: A29316; MUID:87319677; PMID:3498216
 A;Accession: A29316
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 'S',280-481,786-1064 <HUR>
 A;Cross-references: GB:M17421; NID:g161474; PIDN:AAA30050.1; PID:g552260
 R;Hunt, L.T.; Barker, W.C.
 FASEB J. 3, 1760-1764, 1989
 A;Title: Avidin-like domain in an epidermal growth factor homolog from a sea
 urchin.
 A;Reference number: A43131; MUID:89196806; PMID:2784773
 A;Contents: annotation
 C;Comment: EGF homology repeats 10-17 are spliced out in the short form
 (fibropellin Ib).
 C;Superfamily: C1r/C1s repeat homology; EGF homology
 F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-1064/Product: fibropellin I #status predicted <FIB>
 F;23-54/Domain: EGF homology <EG01>
 F;57-175/Domain: C1r/C1s repeat homology <C1R>
 F;180-211/Domain: EGF homology <EG02>
 F;218-249/Domain: EGF homology <EG03>
 F;256-287/Domain: EGF homology <EG04>
 F;294-325/Domain: EGF homology <EG05>
 F;332-363/Domain: EGF homology <EG06>
 F;370-401/Domain: EGF homology <EG07>
 F;408-439/Domain: EGF homology <EG08>
 F;446-477/Domain: EGF homology <EG09>
 F;484-515/Domain: EGF homology <EG10>
 F;522-553/Domain: EGF homology <EG11>
 F;560-591/Domain: EGF homology <EG12>
 F;598-629/Domain: EGF homology <EG13>
 F;636-667/Domain: EGF homology <EG14>
 F;674-705/Domain: EGF homology <EG15>
 F;712-743/Domain: EGF homology <EG16>
 F;750-781/Domain: EGF homology <EG17>
 F;788-819/Domain: EGF homology <EG18>

F;826-857/Domain: EGF homology <EG19>
 F;864-895/Domain: EGF homology <EG20>
 F;902-933/Domain: EGF homology <EG21>
 F;936-1064/Region: avidin-like
 F;23-34,28-43,45-54,62-88,180-191,185-200,202-211,218-229,223-238,240-249,256-267,261-276,278-287,294-305,299-314,316-325,332-343,337-352,354-363,370-381,375-390,392-401,408-419,413-428,430-439,446-457,451-466,468-477,484-495/Disulfide bonds: #status predicted
 F;489-504,506-515,522-533,527-542,544-553,560-571,565-580,582-591,598-609,603-618,620-629,636-647,641-656,658-667,674-685,679-694,696-705,712-723,717-732,734-743,750-761,755-770,772-781,788-799,793-808,810-819,826-837,831-846,848-857,864-875,869-884,886-895,902-913,907-922,924-933/Disulfide bonds: #status predicted

Query Match 14.2%; Score 954.5; DB 2; Length 1064;
 Best Local Similarity 28.0%; Pred. No. 2.5e-43;
 Matches 290; Conservative 93; Mismatches 305; Indels 347; Gaps 69;

```

Qy      30 DPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNWFKCTRHRVSYRTAYRHGEKTMYSR 89
      |||:| :                               :|||:| : ||
Db      181 DPNLCQNG-----AACTDLVNDYACT----- 201

Qy      90 RKSQCCPGFYESGEMC---VPHCA-DKCVHGRCIAPN----TCQCEPGWGGTNCSSACDG 141
      | ||| :| | : || | | :| | | | :| | : :
Db      202 ----CPPGF--TGRNCEIDIDECASDPCQNGGACVDGVNGYVCNCVPGFDGDECENNIN- 254

Qy      142 DHWGPHTSRCQCKNGALCNPITGA----CHCAAGFRGWRCE---DRCEQGTYGNDCHQR 194
      | | | ||| :| : | | | ||| | ||| | |
Db      255 -----ECAS-SPCLNGGIC--VDGVNMFECTCLAGFTGVRCEVNIDECAS----- 296

Qy      195 CQCQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGE 252
      |||| | | : | | || | : || : : | : | || || | :
Db      297 APCQNGGICIDGINGYTCSCPLGFSGDNCEN-----NDDECSS-IPCLNGGTCVDLVNA 349

Qy      253 --CSCPSGWMGTVCQGQPCPEGRFGKNCSEQCQCHNGGTC-DAATG-QCHCSPGYTGERCQ 308
      | | || | | | | : | ||| | | | | |||| | :
Db      350 YMCVCAPGWTGPTCADNIDE-----CA-SAPCQNGGVCIDGVNGYMCDCQPGYTGTTHCE 402

Qy      309 ---DECPVGTYGVLCAETCQCVNNGGKCYH-VSG-ACLCEAGFAGERCEARLCPEGLYGIK 363
      ||| | ||| | | :| :| || | ||
Db      403 TDIDEC-----ARPPCQNGGDCVDGVNGYVCICAPGFDGLNCE----- 440

Qy      364 CDKRCPCHLLENTHSC--HPMSGECACKPGWSGLYCNETCSPGFYGEACQ---QICS---C 415
      | | | | :| | ||| | : | : | : |
Db      441 -----NNIDECASRPCQNGAVCVDGVNGFVC--TCSAGYTGVLCETDINECASMPIC 489

Qy      416 QNGADC-DSVTGK-CTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVAVCS-PVDG-SC 471
      || | | | | |||| ||:| :| | :| | | :| | |
Db      490 LGGVCTDLVNGYICTCAAGFEGTNCETDTD-----ECAS-FPCQNGATCTDQVNGYVC 542

Qy      472 TCKAGWHGVDC-----SIRCPGSG-----TWGFGCNL-----TCQ----- 500
      || | : || | | | :| :| | : || :
Db      543 TCVPGYTGVLCETDINECASFPCLNGGTCNDQVNGYVCVCAQDTSVSTCETDRDECASAP 602

Qy      501 CLNGGAC-NTLDG-TCTCAPGWRGEKCEL-----PCQDGTGYNCAERCDCSHADGC 550
      ||||| : : :| ||| ||| | || : || :| || | : : | :
Db      603 CLNGGACMDVNGFVCTCLPGWEGTNCIEINTDECASSPCMNG--GL-CVDQVN-SYV--- 655

```

Qy 551 HPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASC--SPDDGICECAPGFRG 608
 | | | | : : | : | | | | :
 Db 656 -----CFCLPGFTGIHCGTEIDECASSP-----CLNGGQCIDRVDSYECVCAAGYTA 702

Qy 609 TTCQ-----RICSPGFYGHRCSTCPQCVHSSGPC----- 638
 | | : : | : | : | : | :
 Db 703 VRCQINIDECASAPCQNGGVCVDGVNGYVCNCAPGYTGDNCETEIDEC--ASMPCLNGGA 760

Qy 639 --HHITG-LCDCLPGFTGALC-----NEVCPSGRFGKNCAGICTCTNNGTCNPIDRS 687
 : | | : : | : | : | : | :
 Db 761 CIEMVNGYTCQCVAGYTGVICETDIDECASAPCQNG-----GVCTDTINGYI----- 807

Qy 688 CQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDG---ECKCTPGWTGLYCT 743
 | | | : | : | : | : | : | : | :
 Db 808 CACVPGFTGSNCETNIDECASDP-----CLNGGIC--VDGVNGFVCQCPNYSPTYCE 858

Qy 744 QRCPLGFYGKDCALICQCQNGADCDHISGQ--CTCRTGFMGRHCE---QKCPSTYGYGC 798
 | | | | | : : | : | : : | : :
 Db 859 -----ISLDACRSMPCQNGATCVNVGADYVCECVPGYAGQNCIDINECAS----- 904

Qy 799 RQICDCLNNSTC-DHITG-TCYCSPGWKGARCDQAGVIVGNL-----NSLSRTSTA 848
 | | | | | | : | : : | : : : :
 Db 905 ---LPCQNGGLCIDGIAGYTCQRLGYIGVNCEEVGFCDLEGMWYNECNDQVTITKTSTG 961

Qy 849 LPADSYQIGAIAGIIILVLVVLFLALFIIYRHKQKGKESMPAVTYTPAMRVVNADYTI 908
 : : : : | : : | : | : | :
 Db 962 M-----MLGDYMTYNERALGYAAPTVVVGYASN----NYDFPS 995

Qy 909 SG-TLPHSNGGNANS 922
 | : | : | :
 Db 996 FGFTVVRDNGQSTTS 1010

RESULT 15

T09059

notch4 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 08-Sep-2002

C;Accession: T09059

R;Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.;

Loretz, C.; Schmidt, S.; Tipton, S.; Traicoff, R.; Zackrone, K.; Hood, L.

submitted to the EMBL Data Library, October 1997

A;Description: Sequence of the mouse major histocompatibility locus class III region.

A;Reference number: Z16543

A;Accession: T09059

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1964 <ROW>

A;Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564947

C;Genetics:

A;Gene: notch4

A;Map position: 17

A;Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 671/2; 720/1; 771/1; 810/2; 839/3; 890/1; 951/3; 1036/1; 1073/3; 1248/2; 1376/2; 1435/1; 1508/2; 1531/3; 1578/1; 1679/3; 1729/1; 1761/3

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

C;Keywords: receptor; signal transduction
F;514-545/Domain: EGF homology <EGF>

Query Match 14.1%; Score 952.5; DB 2; Length 1964;
Best Local Similarity 26.5%; Pred. No. 5.4e-43;
Matches 298; Conservative 68; Mismatches 315; Indels 445; Gaps 65;

```
Qy      95 CP-GFYESGEMCVPHCADKCV-----HGRCIAPNT---CQCEPGWGGTNCSSACDGDH 143
      || || :|: | | : | | | : | |||| | |
Db     102 CPSGF--TGDRCQTHLEELCPPSFCSNNGGHCVYQASGRPQCSCEPGWTGEQCQLR----- 154

Qy     144 WGPHTSRCQCKNGALCNPITG--ACHCAAGFRGWRCE-----DRCEQGTYGND 191
      |: | || |: | | | | | | | | | | | |
Db     155 --DFCSAN-PCANGGVCLATYPQIQCRCPGFEGHTCERDINECFLEPGPCPQGT--SC 208

Qy     192 HQ-----RCQ-----CQNGATCD-----HVTGE-CRCPPGYTGA 219
      | :|: | | | | | | | | | | | |
Db     209 HNTLGSYQCLCPVGQEGPQCKLRKGACPPGSCLNGGTCQLVPEGHSTFHLCLCPPGFTGL 268

Qy     220 FCE-----DLCPGPKHG-----PQCEQRCP--CQNGGV 245
      || | | | | | | | | | | | |
Db     269 DCEMNPDDCVRHQCQNGATCLDGLDITYTCLCPKTKWGWDCSEDI DECEARGPPRCRNGGT 328

Qy     246 CHHVTG--ECSCPSGWMGTVCQP-----CPEGRFGKNCS 278
      | : | | | | | | | | | | | |
Db     329 CQNTAGSFHCVCVSGWGAGCEENLDDCAAATCAPGSTCIDRVGSFSCLCPPGRTGLLCH 388

Qy     279 QE--C---QCHNGGTC--DAATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQC 326
      | | | | | : | | | | | | | | | |
Db     389 LEDMCLSQPCHVNAQCSTNPLTGSTLCICQPGYSGSTCHQDLDECQMAQQG-----PSPC 443

Qy     327 VNGGKCYHVSGA--CLCEAGFAGERCEAR-----LCPEGLY 360
      :|| | : | | | | : | | | | | | |
Db     444 EHGGSCINTPGSFNCLCLPGYTGSRCEADHNECLSQPCHPGSTCLDLLATFHCLCPPGLE 503

Qy     361 GIKCD---KRC---PCHLENTHSCHPMMSG--ECACKPGWSGLYCNE----- 398
      | |: | | | | :|| : : | | | | : |
Db     504 GRLCEVEVNECTSNPC--LNQAACHDLLNGFQCLCLPGFTGARCEKDMDECSSTPCANGG 561

Qy     399 -----TCSPGFYGEACQQICS-----CQNGADCDSVTGK--CTCAPGFKGIDC 439
      | ||| | | :|: | | | | : | | | | |
Db     562 RCRDQPGAFYCECLPGFEGPHCEKEVDECLSDPCPVGASCLDLPGAFFCLCRPGFTGQLC 621

Qy     440 STP-----CPLGTYGI-----NCSSRCG----- 457
      | | | | | | | | | | | |
Db     622 EVPLCTPNMCQPGQQCQGEHRAPCLCPDGS PGCVPAEDNCPCHHGHCRSLCVCDEGWT 681

Qy     458 -----CKNDAVCSPVDG--SCTCKAGWHGVDCS---IRCPSGTW----- 491
      | : | | : || | | : | | | | |
Db     682 GPECETELGGCISTPCAHHGTCHPQPSGYNCTCPAGYMGLTCSEEVTAHSGPCLNGGSC 741

Qy     492 -----GFGCN-----LTCQCLNGGACNTLDGT--CTCAPGWRGEKCE- 526
      |: | : | | | | | | | | | | |
Db     742 SIRPEGYSCTCLPSHTGRHCQTAVDHCVSASCLNGGTCVNKPGTFFCLCATGFQGLHCEE 801

Qy     527 -----LPCQDGTYGLNC-----AERCDCSHADGCHPT-- 553
      ||| | | | | | | | | | :
```

Db 802 KTNPSCADSPCRNKATCQDTPRGARCLCSPGYTGSSCQTLIDLCLARKPCPHRTARCLQSGP 861
 Qy 554 TGHCRCLPGWSGVHCD--SVCAEGRWGPNCSLPCYCKNGASCSPDDG---ICECAPGFRG 608
 : | | | | : | | : : : | | | | : |
 Db 862 SFQCLCLQGTGALCDFPLSCQKAAMSQGIETISGLCQNGGLCI-DTGSSYFCRCPPGFQG 920
 Qy 609 TTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKN 668
 | | : | | | | : | : | : | | | : |
 Db 921 KLCQDNVNP-----C--EPNPGHGS---TCVPQPSGYVCQ--CAPGYEQN 960
 Qy 669 CAGI---C---TCTNNGTC--NPIDRSCQCPGWIGSDC-----SQPCPPAHWGPNC 712
 | : : | | : | | | | | | : | | : |
 Db 961 CSKVLDACQSQPCHNHGTCTSRPGGFHCACPPGFVGLRCEGDVDECLDRPCHPS----- 1014
 Qy 713 IHTCNCHNGAFCSAYDGECKCTPGWTGLYC-----TQ 744
 | | : : : | : | | | | | | |
 Db 1015 -GTAACH--SLANAF--YCQCLPGHTGQRCVEMDLCSQPCSNGGSCEITTGPPPGFTC 1069
 Qy 745 RCPLGFYKDC---ALIC---QCQNGADC-----DHISGQCTCRTGFMGRHC-EQKCPSPG 792
 | | | | | | | | | | | | : | | | | |
 Db 1070 HCPKGFEGPTCSHKALSCGIHHCHNGGLCLPSPKPGSPPLCACLSGFGGPDCLTPPAPP- 1128
 Qy 793 TYGYGCRQICDCLNNSTCDHITG-----TCYCSPGWKGARCDQAG 832
 | | | : | | | | | | | : |
 Db 1129 ----GCGPPSPCLHNGTCTETPGLGNPGFQCTCPPDSPGPRCQRP 1170

Search completed: March 26, 2004, 16:12:06
 Job time : 36.117 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:11:16 ; Search time 52.8389 Seconds
(without alignments)
5645.353 Million cell updates/sec

Title: US-10-092-390-2
Perfect score: 6744
Sequence: 1 MVISLNSCLSFICLLCHWI.....SSPKQEDSGGSSSNSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result		%	Query				
No.	Score	Match	Length	DB	ID		Description

1	6744	100.0	1140	14	US-10-092-390-2	Sequence 2, Appli
2	6744	100.0	1140	15	US-10-052-648A-33	Sequence 33, Appl
3	3769	55.9	969	15	US-10-052-648A-34	Sequence 34, Appl
4	3603	53.4	969	15	US-10-052-648A-35	Sequence 35, Appl
5	3601	53.4	586	14	US-10-092-390-4	Sequence 4, Appli
6	3399	50.4	601	14	US-10-365-227-20	Sequence 20, Appl
7	2687	39.8	1037	15	US-10-052-648A-10	Sequence 10, Appl
8	2668	39.6	1034	15	US-10-052-648A-31	Sequence 31, Appl
9	2667	39.5	1034	15	US-10-052-648A-32	Sequence 32, Appl
10	2656	39.4	1037	15	US-10-052-648A-8	Sequence 8, Appli
11	2506.5	37.2	1050	10	US-09-796-753-114	Sequence 114, App
12	2472.5	36.7	928	15	US-10-052-648A-4	Sequence 4, Appli
13	2472.5	36.7	928	15	US-10-052-648A-6	Sequence 6, Appli
14	2451.5	36.4	1020	15	US-10-052-648A-2	Sequence 2, Appli
15	1909	28.3	636	10	US-09-796-753-100	Sequence 100, App
16	1909	28.3	636	10	US-09-796-753-124	Sequence 124, App
17	1881.5	27.9	1393	12	US-10-312-352-21	Sequence 21, Appl
18	1847.5	27.4	739	12	US-10-211-462-69	Sequence 69, Appl
19	1805.5	26.8	1111	15	US-10-369-493-5280	Sequence 5280, Ap
20	1522	22.6	384	14	US-10-106-698-6253	Sequence 6253, Ap
21	1192	17.7	474	14	US-10-084-994-11	Sequence 11, Appl
22	1192	17.7	474	14	US-10-193-109-11	Sequence 11, Appl
23	1192	17.7	474	15	US-10-193-409-11	Sequence 11, Appl
24	1044.5	15.5	2556	12	US-10-072-012-134	Sequence 134, App
25	1036	15.4	2524	15	US-10-190-115-25	Sequence 25, Appl
26	1036	15.4	2524	15	US-10-369-072-25	Sequence 25, Appl
27	1034.5	15.3	2444	9	US-09-944-849-2	Sequence 2, Appli
28	1034.5	15.3	2444	12	US-10-072-012-469	Sequence 469, App
29	1034.5	15.3	2556	12	US-10-072-012-467	Sequence 467, App
30	1034.5	15.3	2556	15	US-10-294-006-12	Sequence 12, Appl
31	1032	15.3	2555	12	US-10-072-012-468	Sequence 468, App
32	1024.5	15.2	2447	15	US-10-190-115-28	Sequence 28, Appl
33	1024.5	15.2	2447	15	US-10-369-072-28	Sequence 28, Appl
34	1024	15.2	2531	12	US-10-072-012-470	Sequence 470, App
35	1024	15.2	2531	12	US-10-072-012-471	Sequence 471, App
36	1024	15.2	2531	15	US-10-190-115-29	Sequence 29, Appl
37	1024	15.2	2531	15	US-10-369-072-29	Sequence 29, Appl
38	1014.5	15.0	1473	15	US-10-190-115-4	Sequence 4, Appli
39	1014.5	15.0	1473	15	US-10-369-072-4	Sequence 4, Appli
40	1011	15.0	241	14	US-10-084-994-8	Sequence 8, Appli
41	1011	15.0	241	14	US-10-193-109-8	Sequence 8, Appli
42	1011	15.0	241	15	US-10-193-409-8	Sequence 8, Appli
43	998	14.8	2471	15	US-10-190-115-27	Sequence 27, Appl
44	998	14.8	2471	15	US-10-369-072-27	Sequence 27, Appl
45	980	14.5	2469	15	US-10-190-115-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
 US-10-092-390-2
 ; Sequence 2, Application US/10092390
 ; Publication No. US20030013865A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yu, Xuanchuan

```

; APPLICANT: Miranda, Maricar
; TITLE OF INVENTION: No. US20030013865A1el Human EGF-Family Proteins and
Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1140
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-2

```

```

Query Match          100.0%; Score 6744; DB 14; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

Qy     61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

Qy    121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED 180

Qy    181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPC 240

Qy    241 QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP 300

Qy    301 GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGACLCEAGFAGERCEARLCPEGLY 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGACLCEAGFAGERCEARLCPEGLY 360

Qy    361 GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420

Qy    421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV 480

Qy    481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 DCSIRCPSGTWGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE 540

Qy    541 RCDCSHADGCHPTTGHCRCPLGWSGVHCDVCAEGRWGPNCSLPCYCKNGASCSPDDGIC 600

```

Db	541		RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNC SLPCYCKNGASCSPDDGIC	600
Qy	601		ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSSGPCHHITGLCDCLPGFTGALCNEVC	660
Db	601		ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSSGPCHHITGLCDCLPGFTGALCNEVC	660
Qy	661		PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Db	661		PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Qy	721		GAFC SAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF	780
Db	721		GAFC SAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF	780
Qy	781		MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGV IIVGNLN	840
Db	781		MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGV IIVGNLN	840
Qy	841		SLSRTSTALPADSYQIGAIAGI I I I LVLVVLFL LALFIIYRHKQKGKESSMPAVTYTPAMR	900
Db	841		SLSRTSTALPADSYQIGAIAGI I I I LVLVVLFL LALFIIYRHKQKGKESSMPAVTYTPAMR	900
Qy	901		VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVT KSKNNQLFV	960
Db	901		VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVT KSKNNQLFV	960
Qy	961		NLKNVNP GKRGPVG DCTGTLPADWKHGGYLNELGAFGLDRSYM GKSLKDLGKNSEYNSSN	1020
Db	961		NLKNVNP GKRGPVG DCTGTLPADWKHGGYLNELGAFGLDRSYM GKSLKDLGKNSEYNSSN	1020
Qy	1021		CSLSSSEN PYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT	1080
Db	1021		CSLSSSEN PYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT	1080
Qy	1081		VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1140
Db	1081		VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1140

RESULT 2

US-10-052-648A-33

; Sequence 33, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera

```

; APPLICANT: Rothenberg, Mark
; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 1140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-33

```

```

Query Match          100.0%; Score 6744; DB 15; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60
        |||||||
Db      1 MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC 60

Qy     61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120
        |||||||
Db     61 TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA 120

Qy    121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE 180
        |||||||
Db    121 PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE 180

```

Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Qy	361	GIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Db	361	GIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGV	480
Qy	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAE	540
Db	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAE	540
Qy	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSRDDGIC	600
Db	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSRDDGIC	600
Qy	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVC	660
Db	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVC	660
Qy	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Db	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Qy	721	GAFC SAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF	780
Db	721	GAFC SAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGF	780
Qy	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLN	840
Db	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLN	840
Qy	841	SLSRTSTALPADSYQIGAIAGIILVLVVLFLALFLIYRHKQKGKESSMPAVTYTPAMR	900
Db	841	SLSRTSTALPADSYQIGAIAGIILVLVVLFLALFLIYRHKQKGKESSMPAVTYTPAMR	900
Qy	901	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATS PHVNNRDRMTVTKSKNNQLFV	960
Db	901	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATS PHVNNRDRMTVTKSKNNQLFV	960
Qy	961	NLKNVNPGRGPVGDCTGTL PADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN	1020
Db	961	NLKNVNPGRGPVGDCTGTL PADWKHGGYLNELGAFGLDRSYMGKSLKDLGKNSEYNSSN	1020
Qy	1021	CSLSSEN PYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT	1080

Db	1021	CSLSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPT	1080
Qy	1081	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1140
Db	1081	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1140

RESULT 3

US-10-052-648A-34

; Sequence 34, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Colman, Steven

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Gerlach, Valerie

; APPLICANT: Gunther, Erik

; APPLICANT: Kekuda, Ramesh

; APPLICANT: MacDougall, John R.

; APPLICANT: Mehraban, Fuad

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Shimkets, Richard

; APPLICANT: Smithson, Glennnda

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Stone, David J.

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/284,549

; PRIOR FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: 60/303,229

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,605

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/269,098

; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: 60/264,159

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/265,517

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/271,855

; PRIOR FILING DATE: 2001-02-27

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-34

Query Match 55.9%; Score 3769; DB 15; Length 969;
Best Local Similarity 58.6%; Pred. No. 1.2e-241;
Matches 600; Conservative 126; Mismatches 208; Indels 90; Gaps 4;

```
Qy      109 CADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACH 168
      | ::|||||::|:| | | | | | :| | | | | | | | | | | | | | | | | | | | | | |
Db      28 CTEECVHGRCVSPDTCHCEPGWGGPDCSSGCDSDHWGPHCSNRCQCNQALCNPITGACV 87

Qy      169 CAAGFRGWRCEDRCEQGTGYNDCHQRCQCNQATCDHVTGECRCPPGYTGAFCEDLCPPG 228
      | | | | | | | | | | : | | | : | | | | | | | | | | | | | | | | | | | |
Db      88 CAAGFRGWRCEELCAPGTHGKGCQLPCQCRHGASCDPRAGECLCAPGYTGVYCEELCPPG 147

Qy      229 KHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGT 288
      | | | | | | | | | | | | : | | | : | | | | | | | | | | | | | | | | | |
Db      148 SHGAHCELRCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSDCPCHHGGQ 207

Qy      289 CDAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGKCYHVSACLCCEAGFAGE 348
      | | | | | | | | : | | | : | | | | | | : | | : | | | | | | | | | |
Db      208 CDHVTGQCHCTAGYMGDRQEECPFGSFGFQCSQRCDCCHNGGQCSPTTGACECEPGYKGP 267

Qy      349 RCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEA 408
      | | : | | | | | : | | | | | | : | | | | | | | | | | | | | | | | |
Db      268 RCQERLCPEGLHGPGLTLPCCDADNTISCHPVGTGACTCQPGWSGHHHCNESCVPVGYGDDG 327

Qy      409 CQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVD 468
      | | | : | | | | | | : | | | | | | | | : | | | | | | | | | | | | |
Db      328 CQLPCTCQNGADCHSITGGCTCAPGFMGEVCAVSCAAGTYGPNCSSICSCNNGGTCSVPD 387

Qy      469 GSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELP 528
      | | | | | | | | | | : | | | : | | | | | | | | : | | | : | | | | | |
Db      388 GSCTCKEGWQGLDCTLPSPGTWGLNCNESCTCANGAACSPIDGSCSCTPGWLGDTCCELP 447

Qy      529 CQDGTYGLNCAERCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCYCK 588
      | | | | : | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db      448 CPDGTFTGLNCSEHCDCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNCVSVSCSE 507

Qy      589 NGASCSPDDGICECAPGFRGTTTCQRICSPGFYGHRCSTQCPQCVHSSGPCHHITGLCDCL 648
      | | | | | | : | | | | | | | | | | | | | | | | | : | | | : | | | |
Db      508 NGGSCSPEDGSCECAPGFRGPLCQRICPPGFYGHGCAQPCPLCVHSSRPCHHISGICECL 567

Qy      649 PGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHW 708
      | | | : | | | | : | | | : | | | | | | : | | | | | | | | | | | |
Db      568 PGFSGALCNQVCAGGYFGQDCAQLCSCANNGTCSPIDGSCQCFPGWIGKDCSQACPPGEW 627

Qy      709 GPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCNQADCD 768
      | | | | | | : | | | | | | | | | | | | | | | | | : | | | | | |
Db      628 GPACFHACSCHNGASCSAEDGACHCTPGWTGLFCTQRCPAAFFGKDCGRVCQCNQASCD 687
```

Qy 769 HISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARC 828
 |||:||||| | :|||:| ||:||||:|:|:|:|||||:|||||:| ||
 Db 688 HISGKCTCRTGFTGQHCEQRCAPGTFGYGCQQLCECMNNSTCDHVTGTCTCYCSPGFKGIRC 747
 Qy 829 DQAGVITVGNLNSLSRTSTALPADSYQIGAIAGIILVLVLLFLLALFIIYRHKQKGKES 888
 ||| ::| || :| || |: :||: ||::| :| || :| :||
 Db 748 DQA-ALMMEELNPYTKISPALGAERHSVGAVTGIMLLLFFIVVLLGLFAWHRRRQKEKGR 806
 Qy 889 SM-PAVTYTPAMRVVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDR 947
 :| |:|||||: :|:|
 Db 807 DLAPRVSYTPAMRMTSTDYSL----- 828
 Qy 948 MTVTKSKNNQLFVNLKNVNPGKRGVPGDCTGTLPADWKHGGYLNELGAFGLDRS----YM 1003
 || |:| |
 Db 829 -----GACGMDDRRQNTYIM 842
 Qy 1004 GKSLKDLGKNSEYNSSNCSLSSSENPHYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAE 1063
 | || | :|| ||:|||||:| | | ||||| ||| :
 Db 843 DKGFKDYMKESVCSSTCSLNSSENPHYATIKDPPILTCKLPESYVEMKSPVHMGSPYTD 902
 Qy 1064 INNSTSANRNVYEVEPTVSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSP 1123
 : : ::|:|:|||||:| :| |: |||:|||| | ||||| | :
 Db 903 VPSSLSTSNKNIYEVEPTVSVVQEGCGHNSSYIQNAYDLPRNSHIPGHYDLLPVRQSPANG 962
 Qy 1124 KQED 1127
 :|
 Db 963 PSQD 966

RESULT 4

US-10-052-648A-35

; Sequence 35, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Smithson, Glennnda
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Stone, David J.
 ; APPLICANT: Vernet, Corine A.M.
 ; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

```

; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
;   LENGTH: 969
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (848)..(889)
;   OTHER INFORMATION: Where Xaa is any amino acid
US-10-052-648A-35

```

Query Match 53.4%; Score 3603; DB 15; Length 969;
Best Local Similarity 56.0%; Pred. No. 1.2e-230;
Matches 573; Conservative 124; Mismatches 237; Indels 90; Gaps 4;

Qy	109	CADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACH	168
Db	28	CTEECVHGRCSVSPDTCHEPGWGGPDCSSGCDSDHWGPHCSNRCQCQNGALCNPITGACV	87
Qy	169	CAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPPG	228
Db	88	CAAGFRGWRCEELCAPGTHGKGCQLPCQCRHGASCDPRAGECLCAPGYTGVCYCEELCPPG	147
Qy	229	KHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQQPCPEGRFGKNCSQECQCHNGGT	288
Db	148	SHGAHCELRCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSDQDCPCHHGGQ	207
Qy	289	CDAATGQCHCSPGYTGERCQDECPVGTYGVLCETCQCVNNGGKCYHVSAGACLCEAGFAGE	348
Db	208	CDHVTGQCHCTAGYMGDRCEECPPGSGFGFQCSQHCDCHNGGQCSPTTGACECEPGYKGP	267
Qy	349	RCEARLCPEGLYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEA	408
Db	268	RCQERLCPEGLHGPGLTLPCCDADNTISCHPVTGACTCQPGWSGHHCNESCPVGYGDDG	327

Qy	409	CQQICSCQNGADCDSVTGKCTCAPGFGKIDCSTPCPLGTYGINCSSRCGCKNDAVCS	468
Db	328	CQLPCTCQNGADCHSITGGCTCAPGFMGEVCAVSCAAGTYGPNCSSITCSNNGGTCSPVD	387
Qy	469	GSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELP	528
Db	388	GSCTCKEGWQGLDCTLPCPSGTWGLNCNESCTCANGAACSPIDGSCSCTPGWLGDTCCLP	447
Qy	529	CQDGTYGLNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCK	588
Db	448	CPDGTFTGLNCSEHCDCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNCSSVCSCE	507
Qy	589	NGASCSRDDGICECAPGFRGTTTCQRICSPGFYGHRCSTQCPQCVHSSGPCHHITGLCDCL	648
Db	508	NGGSCSPEDGSCECAPGFRGPLCQRICPPGFYGHGCAQPCPLCVHSSRPCHHISGICECL	567
Qy	649	PGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHW	708
Db	568	PGFSGALCNQVCAGGYFGQDCAQLCSCANNGTCSPIDGSCQCFPGWIGKDCSQACPPGFW	627
Qy	709	GPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCD	768
Db	628	GPACFHACSCHNGASCSAEDGACHCTPGWTGLFCTQRCPAAFFGKDCGRVCQCQNGASCD	687
Qy	769	HISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTTCYCSPGWKGARC	828
Db	688	HISGKCTCRTGFTGQHCEQRCAPGTFGYGCQQLCECMNNSTCDHVTGTTCYCSPGFKGIRC	747
Qy	829	DQAGVIVGNLNSLSRTSTALPADSYQIGAIAGIIILVLVFLFLLALFIIYRHKQKGKES	888
Db	748	DQA-ALMMEELNPYTKISPALGAERHSVGAVTGIMLLLFLIVVLLGLFAWHRRRQKEKGR	806
Qy	889	SM-PAVTYTPAMRVVNADYTIISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDR	947
Db	807	DLAPRVSYTPAMRMTSTDYSLS-----	828
Qy	948	MTVTKSKNNQLFVNLKNNVNPGRGPVGDCGTGLPADWKHGGYLNELGAFGLDRS-----YM	1003
Db	829	-----GACGMDDRRQNTYIM	842
Qy	1004	GKSLKDLGKNSEYNSSNCSLSSSENPHYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAE	1063
Db	843	DKGFKXXMKSPPVHMGSPYTD	902
Qy	1064	INNSTSANRNVYEVEPTVSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSP	1123
Db	903	VPSLSTSNKNIYEVEPTVSVVQEGCGHNSSYIQNAYDLPRNSHIPGHYDLLPVRQSPANG	962
Qy	1124	KQED 1127	
Db	963	PSQD 966	

RESULT 5
US-10-092-390-4
; Sequence 4, Application US/10092390

```
; Publication No. US20030013865A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; TITLE OF INVENTION: No. US20030013865A1el Human EGF-Family Proteins and
Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 586
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-4
```

Query Match 53.4%; Score 3601; DB 14; Length 586;
Best Local Similarity 100.0%; Pred. No. 9.3e-231;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVISLNSCLSFICLLLC HWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	1	MVISLNSCLSFICLLLC HWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTM YRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	61	TDILNWFKCTRHRVSYRTAYRHGEKTM YRRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Qy	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE	180
Db	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE	180
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCP PGKHGPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCP PGKHGPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVC GQPCPEGRFGKNCSQECQCHNGGTCDAA TGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVC GQPCPEGRFGKNCSQECQCHNGGTCDAA TGQCHCSP	300
Qy	301	GYTGERCQDECPVGT YGVLCAETCQCVN GGGKCYHVSGACLCEAGFAGERCEARLCPEG LY	360
Db	301	GYTGERCQDECPVGT YGVLCAETCQCVN GGGKCYHVSGACLCEAGFAGERCEARLCPEG LY	360
Qy	361	GIKCDKRCPC HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQ QICSCQNGAD	420
Db	361	GIKCDKRCPC HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQ QICSCQNGAD	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGT YGINCSSRCGCKNDAVCS PVDGSCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGT YGINCSSRCGCKNDAVCS PVDGSCTCKAGWHGV	480
Qy	481	DCSIRCPSGTWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPCQDGT YGLNCAE	540

Qy	526	ELPCQDGTYGLNCAERCDCHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPC	585
Db	121	ELPCQDGTYGLNCAERCDCHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPC	180
Qy	586	YCKNGASCSPDDGICECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHITGLC	645
Db	181	YCKNGASCSPDDGICECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHITGLC	240
Qy	646	DCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPP	705
Db	241	DCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPP	300
Qy	706	AHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGA	765
Db	301	AHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGA	360
Qy	766	DCDHISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKG	825
Db	361	DCDHISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKG	420
Qy	826	ARCDQAGVVIIVGNLNSLSRTSTALPADSYQIGAIAGIIILVLVFLFLLALFIIYRHKQKG	885
Db	421	ARCDQAGVVIIVGNLNSLSRTSTALPADSYHIGAIAGIIILVLVFLFLLALFIIYRHKQKG	480
Qy	886	KESSMPAVTYTPAMRVVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNR	945
Db	481	KESSMPAVTYTPAMRVVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNR	540
Qy	946	DRMTVTKSKNNQLFVNLKNVNPGRGPVGDCGTG	978
Db	541	DRMTVTKSKNNQLFVNLKNVNPGRGPVGDCMG	573

RESULT 7

US-10-052-648A-10

; Sequence 10, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Smithson, Glennda
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Stone, David J.
 ; APPLICANT: Vernet, Corine A.M.


```

; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-10

```

```

Query Match          39.8%; Score 2687; DB 15; Length 1037;
Best Local Similarity 43.6%; Pred. No. 8e-170;
Matches 506; Conservative 115; Mismatches 361; Indels 178; Gaps 27;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCSTDILNW---FKCT 70
      ||| : | || ||| |||::|::|: || : | | |
Db      9 LLLAVGLRLAGTLNPSDENTCSFWESFTTTTKESHSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || ||::|: ||| ||||| | ||| || :|||::||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQRLQCCHGFYESREFCVPLCAQECVHGRCVAPNQCQCVPGW 126

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | : |:| :| | :| : | | |
Db     127 RGDDCSSECAPGMWGPQCDKPKSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCTPGYYGPA 186

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVT 250
      | |||| :|| || || | || || : | | | | |||||
Db     187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHPCQNGGVFQTPQ 245

Qy     251 GECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAATGQCHCSPGYTGERCQDE 310
      | |||| |||||:| |||| | |||||:|||| | | |||| |:|||||:|::|

```

Db 246 GSCSCPPGWMGTICSLPCPEGFHGPNCSSQECRCHNGGLCDRFTGQCRCAPGYTGDRCREE 305
 Qy 311 CPVGTYGVLC AETCQCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
 |||| :| |||| | :| :||| | | :| | ||| :| || :| |
 Db 306 CPVGRFGQDCAETCDCAPDARCFPANGACLCEHGFTGDRCTDRLCPDGFYGLSCQAPCTC 365
 Qy 371 HLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
 | :| |||| :||| :| ||| :||| :| :| || :| | :| | :| :| |
 Db 366 DREHSLSCHPMNGECSCLPGWAGLHCNESCPODTHGPGCQEYCLCLHGGVCQATSGLCQC 425
 Qy 431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGWHGVDCSIRCPSTG 490
 ||| :| | :| || ||| :||| :| | ||| :| | ||| :||| :| |
 Db 426 APGYTGPHCASLCPDPTYGVNCSARCSCEAIACSPIDGECVCKEGWQRGNC SVPCPPGT 485
 Qy 491 WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAERCDCSHADGC 550
 ||| || :||| : | :| ||| ||| | :||| | :| || |||| :|||
 Db 486 WGFSCNASQCQAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASRCDCHSDGC 545
 Qy 551 HPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTT 610
 | | | :| | | | | | | | | | | | :| | :| | | | | :
 Db 546 DPVHGRCQCQAGWMGARCHLSCEGLWGVNCSNTCTCKNGGTCLPENGNCVCAPGFRGPS 605
 Qy 611 CQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA 670
 ||| | || || | :|
 Db 606 CQRSCQPGRYGKR-----CVP----- 621
 Qy 671 GICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGE 730
 | | | : | :| :| | | | | | | | | | | | | :| | |
 Db 622 --CKCANHSFCHPSNGTCYCLAGWTGPDSCQPCPPGHWGENCAQTCCQCHGGTCHPQDGS 679
 Qy 731 CKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADC DHISGQCTC RTGFMGRHCEQKCP 790
 | | |||| :| : |||| :| :| : ||| | | | | | | | |
 Db 680 CICPLGWTGHHCLEGCPLGTFGANCSQPCQCGPGEKC-----HPE---- 719
 Qy 791 SGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTALP 850
 || | | || || | :| : | :|
 Db 720 -----TGACVCPPGHSGAPCR-----IG----IQEPFTVMP 746
 Qy 851 AD--SY-QIGAIAGIIILVLVVLFLIALFIIYRHKQKGKESMPAVTYTPAMRVVNADYT 907
 :| :||| :| :| :| :||| ||| |||| | | :| :| :|
 Db 747 TTPVAYNSLGAVIGIAVLGSLVVALVALFIGYRHWQKGKEHHHLAVAYSSG-RLDGSEYV 805
 Qy 908 ISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLEFVN LKN-VN 966
 : | : ||| :||| :| :| :| :| :| :| :| :| :| :|
 Db 806 MPDVFP-----SYSHYYSNPSYHTLSQCSNP PPPPNK-----VPGPLFASLQNP ER 851
 Qy 967 PGKRGPVG-DCTGTL PADWKH-----GGYLNELGAFGLDRSYM GKSL-----KDLGK 1012
 || | | | ||||| | :| :| :| |||| | |
 Db 852 PG--GAQGHDNHTTLPADWKHRRPPPGPLDR-GSSHLDRSYSYSYSGPGPFYDKGLIS 908
 Qy 1013 NSEYNSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSP-----ARR 1057
 | :| || ||||| :| | | | :||| | |
 Db 909 EEELGASVTSLSSENPYATIRDLPSLPGGPRESSYMEMKGPPSGSPPRQPPQFWD SQRR 967
 Qy 1058 DSPYAEINNSTSANRNVYEVEPTVSVVQGVFSNNGRLSQDP-----YDLPKN SHIP 1108
 | : :| | | :| :| :| :| | | | | | |
 Db 968 RQPQPQRDSGT-----YE-QPSPL-----IHDRDSVGSQPPLPPLPGHYDSPKN SHIP 1016

Qy 1109 CHYDLLPVRDSSSSP-KQED 1127
| | | | | | | | : : : |
Db 1017 GHYDLPPVRHPPSPPLRRQD 1036

RESULT 8

US-10-052-648A-31

; Sequence 31, Application US/10052648A
; Publication No. US20040005558A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steven
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gunther, Erik
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacDougall, John R.
; APPLICANT: Mehraban, Fuad
; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-052-648A-31

Query Match 39.6%; Score 2668; DB 15; Length 1034;
Best Local Similarity 42.7%; Pred. No. 1.5e-168;
Matches 493; Conservative 110; Mismatches 379; Indels 172; Gaps 16;

Qy 14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNW---FKCT 70
||| : || |||||: |||: | :||| : || : || |
Db 7 LLLALGLRLTGTILNSNDPNVCTFWESFTTTTKESHLRPFSLIPAESCH--RPWEDPHTCA 64

Qy 71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
: | ||| || | : ||| |:||| ||| || :|||:| || ||| |||
Db 65 QPTVVYRTVYRQVVKMDSRPLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW 124

Qy 131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
| :||| | ||| | | | : | :| :||| | :| : | | |||
Db 125 RGGDCSSSECAPGMWGPQCDKFCHCGNNSCDPKSGACFCPSGLQPPNCLQPCPAGHYGPA 184

Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
| ||| ||:| | |||| | | | | : |||||
Db 185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGDGFFCPRTPCQNGGVPQGSQ 243

Qy 251 GECSCPSGWMGTVCGQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
| |||| |||| :| |||| | ||:|:|:| || |||||:| || :||:|
Db 244 GSCSCPPGWMGVICSLPCPEGFHGPNTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

Qy 311 CPVGTGYVLCAETCQCVNNGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKDKRCPC 370
||| :| |||| | | :| : |||| | | :| ||| :| || : | |
Db 304 CPVGRFGQDCAETCDCAPGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

Qy 371 HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSDVTGKCTC 430
| : |||| |||:|:|:|:|:| :| || : | | :| | : :| | |
Db 364 DPEHSLSCHPMHGECSQPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGCLCLADSGLCRC 423

Qy 431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDGSCTCKAGWHGVDCSIRCPST 490
|||: | |: || ||||| ||| :| |||:|:| || || :||: || ||
Db 424 APGYTGPHCANLCPDPTYGINCSSRCSCEAIACSPIDGTICKEGWQRGNCSVPCPLGT 483

Qy 491 WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGC 550
||| || :||| : | | : | ||| ||| | :||| | :| || ||| :|||
Db 484 WGFNCNASQCQAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 543

Qy 551 HPTTGHCRCLPGWSGVHCDSDVCAEGRWGPNCSLPCYCKNGASCSRDDGICECAPGFRGTT 610
| | ||| || | | | ||| ||| | |||| :| :| | ||||| :
Db 544 DPVHGQCRCQAGWMGTRCHLPCPEGFWGANCNTCTCKNGGTCVSENGNCVCAPGFRGPS 603

Qy 611 CQRICSPGFYGHRCSTCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA 670
||| | || || | |
Db 604 CQRPCPPGRYKRCVQ----- 619

Qy 671 GICTCTNN-GTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDG 729
| | || :|:| | :| | || | |||: ||| ||| | | ||:| | ||

Db 620 --CKCNNNHSSCHPSDGTCSCLAGWTGPDCEACPPGHWGLKCSQLCQCHHGGTCHPQDG 677
 Qy 730 ECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKC 789
 | | | | | | | | : | | : | : | : | | | |
 Db 678 SCICTPGWTGPNCLEGCPPRMFGVNCSQLCQCDLG----- 712
 Qy 790 PSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTAL 849
 | | | | | | | | : | : | : | :
 Db 713 -----EMCHPQTGACVCPGHSADCK-----MGSQESFTIMPTS- 747
 Qy 850 PADSQIGAIAGIILVLVLFLLALFIIYRHKQKGESSMPAVTYTPAMRVVNADYTIS 909
 | : | : | | : | : | : | | | | | | | : | : | :
 Db 748 PVTHNSLGAVIGIAVLGTLVVALIALFIGYRQWQKGKEHEHLAVAYSTG-RLDGS DYVMP 806
 Qy 910 GTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDMRTVTKSKNNQLFVNLKNNVPGK 969
 | : | | : | | | | : | : | : | | : | :
 Db 807 DVSP-----SYSHYYSNPSYHTLSQCSNP PPPN-----KVPGSQLFVSSQAPERPS 853
 Qy 970 RGPVGDCTGTLPADWKHGGYLNELGAFGLDRSY-----MGKS 1006
 | : | | | | | | : | | | | | : | |
 Db 854 RAHGRENHVTLPADWKHRRPHERGASHLDRSYSCSYSHRNGPGPFCHKGPIS E E GLGAS 913
 Qy 1007 LKDLGKNSEYNSSNCSLSSENPHYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINN 1066
 : | | | | | | : | | : | | | | | : :
 Db 914 VMSL-----SSENPHYATIRDLPSLPGEPRSEGYVEMKGPPSVSPPRQSLH- 958
 Qy 1067 STSANRNVYEVEP-----TVSVVQGVFSNNGRLSQDP-----YDLPKNSHIPCHY 1111
 : | : : | | : | | | | | | | | | | | |
 Db 959 --LRDRQQRQLQPQRDSGTYEQPSPLSHNEESLGSTPPLPPGLPPGQYDSPKNSHIPGHY 1016
 Qy 1112 DLLPVRDSSSSPKQ 1125
 | | | | | :
 Db 1017 DLPPVRHPPSPPSR 1030

RESULT 9

US-10-052-648A-32

; Sequence 32, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Smithson, Glennnda
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Stone, David J.

```

; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
;   LENGTH: 1034
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-052-648A-32

```

```

Query Match          39.5%;   Score 2667;   DB 15;   Length 1034;
Best Local Similarity 42.8%;   Pred. No. 1.7e-168;
Matches 494;   Conservative 110;   Mismatches 378;   Indels 172;   Gaps 17;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNW---FKCT 70
      ||| : || |||||: |||: | :||: || : || |
Db      7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKESHLRPFSLPAESCH--RPWEDPHTCA 64

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || | | : ||| |:||| ||| || :||| ||| ||| ||| |||
Db      65 QPTVVYRTVYRQVVKMSRPRQLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW 124

Qy      131 GGTNCSSACDGDHWGPHCTSRQCCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | : | : | : | : | : | | |
Db      125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDPKSGTCFCPSGLQPPNCLQPCPAGHYGPA 184

Qy      191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVT 250
      | ||| ||: || | |||| | | | | | : |||||
Db      185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGDFFCPRTPYPCQNGGVPQGSQ 243

Qy      251 GECSCPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310

```

Db	244	GSCSCPPGWMGVICSLPCPEGFHGPNCQTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE	303
Qy	311	CPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLYGIKCDKRCPC	370
Db	304	CPVGRFGQDCAETCDCAPGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC	363
Qy	371	HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC	430
Db	364	DPEHSLSCHPMHGECSQPWAGLHCNESCPQDTHGPGCQEHCLCLHGGLCLADSGLCRC	423
Qy	431	APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPPVDGSCTCKAGWHGVDCSIRCPST	490
Db	424	APGYTGPHCANLCPDPTYGINCSSRCSCEAIACSPIDGTICKEGWQRGNCSVPCPLGT	483
Qy	491	WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAERCDCSHADGC	550
Db	484	WGFCNCASCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC	543
Qy	551	HPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTT	610
Db	544	DPVHGQCRCQAGWMGTRCHLPCPEGFHWGANCSNTCTCKNGGTCVSENGNCVCAPGFRGPS	603
Qy	611	CQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA	670
Db	604	CQRPCPPGRYGKRCVQ-----	619
Qy	671	GICTCTNN-GTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDG	729
Db	620	--CKCNNNHSSCHPSDGTCSCLAGWTGPDCSEACPPGHWGLKCSQLCQCHHGGTCHPQDG	677
Qy	730	ECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGADCDSHISGQCTCRTGFMGRHCEQKC	789
Db	678	SCICTPGWTGPNCLEGCPPRMFGVNCSQLCQCDLGEMC-----HPE----	718
Qy	790	PSGTYGYGCRQICDCLNNSTCDHITGTCTCYSPGWKGARCDQAGVIIVGNLNSLSRTSTAL	849
Db	719	-----TGACVCPPGHSGADCK-----MGSQESFTIMPTS-	747
Qy	850	PADSYQIGAIAGIIILVLVLFLLALFIIYRHKQKGKESMPAVTYTPAMRVVNADYTIS	909
Db	748	PVTHNSLGAVIGIAVLGTLVVALIALFIGYRQWQKGKEHEHLAVAYSTG-RLDGSDYVMP	806
Qy	910	GTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKVNPNPGK	969
Db	807	DVSP-----SYSHYYSNPSYHTLSQCSNPNNPPN-----KVPGSQLFVSSQAPERPS	853
Qy	970	RGPVGDCTGTLPADWKHGGYLNELGAFGLDRSY-----MGKS	1006
Db	854	RAHGRENHVTLPADWKHRRPHERGASHLDRSYSCSYSHRNGPGPFCHKGPISSEGLGAS	913
Qy	1007	LKDLGKNSEYNSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINN	1066
Db	914	VMSL-----SSENPYATIRDPLSLPGEPRESGYVEMKGPPSVSPPRQSLH-	958
Qy	1067	STSANRNVYEVEP-----TVSVVQGVFSNNGRLSQDP-----YDLPKNSHIPCHY	1111

Db 959 --LRDRQQRQLQPQRDSGTYEQPSPLSHNEESLGSTPPLPPLPGLPPGHYDSPKNSHIPGHY 1016

Qy 1112 DLLPVRDSSSSPKQ 1125
 || ||| | | :
 Db 1017 DLPPVRHPPSPPSR 1030

RESULT 10

US-10-052-648A-8

; Sequence 8, Application US/10052648A
 ; Publication No. US20040005558A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Anderson, David
 ; APPLICANT: Burgess, Catherine
 ; APPLICANT: Casman, Stacie
 ; APPLICANT: Colman, Steven
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Mehraban, Fuad
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Smithson, Glennnda
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Stone, David J.
 ; APPLICANT: Vernet, Corine A.M.
 ; APPLICANT: Zerhusen, Bryan D.
 ; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
 ; TITLE OF INVENTION: USING THE SAME
 ; FILE REFERENCE: 21402-250 (CURA-550)
 ; CURRENT APPLICATION NUMBER: US/10/052,648A
 ; CURRENT FILING DATE: 2002-12-09
 ; PRIOR APPLICATION NUMBER: 60/262,454
 ; PRIOR FILING DATE: 2001-01-18
 ; PRIOR APPLICATION NUMBER: 60/272,920
 ; PRIOR FILING DATE: 2001-03-02
 ; PRIOR APPLICATION NUMBER: 60/284,549
 ; PRIOR FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: 60/303,229
 ; PRIOR FILING DATE: 2001-07-05
 ; PRIOR APPLICATION NUMBER: 60/262,892
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: 60/263,605
 ; PRIOR FILING DATE: 2001-01-23
 ; PRIOR APPLICATION NUMBER: 60/269,098
 ; PRIOR FILING DATE: 2001-02-15
 ; PRIOR APPLICATION NUMBER: 60/264,159
 ; PRIOR FILING DATE: 2001-01-25
 ; PRIOR APPLICATION NUMBER: 60/265,517
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/271,855
 ; PRIOR FILING DATE: 2001-02-27
 ; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-8

Query Match 39.4%; Score 2656; DB 15; Length 1037;
Best Local Similarity 43.4%; Pred. No. 9.1e-168;
Matches 503; Conservative 115; Mismatches 364; Indels 178; Gaps 27;

Qy 14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW--FKCT 70
| | | : | | | | | | | | | | : | : | | : | | : | | | |
Db 9 LLLAVGLRLAGTILNPSDPNTCSFWESFTTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy 71 RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
: | | | | | | | | | | : | : | | | | | | | | | | | | | | | |
Db 67 QPTVVYRTVYRQVVKTDHRQRLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCVPGW 126

Qy 131 GGTNCSSACDGDHWGPHCTSRQCCKNGALCNPI TGACHCAAGFRGWRCEDRCEQGTYGND 190
| : | | | | | | | | | | | | : | : | : | | : | : | | | | |
Db 127 RGDDCSSECAPGMWGPQCDKPCSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCTPGYYGPA 186

Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVT 250
| | | | : | | | | | | | | | | : | | | | | | | | | | | | |
Db 187 CQFRCQC-HGAPCDPQTGACFCPAERTGPS CDVSCSQGTSGFFCPSTHPCQNGGVFQTPQ 245

Qy 251 GECSCPSGWMGTVC GQPCPEGRFGKNCSQECQCHNGGT CDAATGQCHCSPGYTGERCQDE 310
| | | | | | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db 246 GSCSCPPGWMGTICSLPCPEGFHGPNC SQECRCHNGGLCDRFTGQCRCAPGYTGDRCREE 305

Qy 311 CPVGTYGVLC AETCQCVNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
| | | | : | | | | | | | | : | : | | | | | | | | | | | | | | |
Db 306 CPVGRFGQDCAETCD CAPDARCFPANGACLCEHGFTGDRCTDR LCPDGFYGLSCQAPRTC 365

Qy 371 HLENTHSCHPMMSGECACKPGW SGLYCNETCS PGFYGEACQQICSCQNGADCD SVTGKCTC 430
| : | | | | : | | | : | | | : | | : | | : | | : | | | | |
Db 366 DREHSLSCHPMNGECSCLPGWAGLHCNESCPQDTHGPGCQEHCLCLHGGVCQATSGLCQC 425

Qy 431 APGEFKGIDCSTPCPLGT YGINCSSRCGCKND AVCS PVDG SCTCKAGWHGVDCSIRCPSGT 490
| | | : | | : | | | | | : | | | | | | | | | | : | | | | |
Db 426 APGYTGPHCASLCP PDTYGVNCSARCS CENAIACSPIDGECVCKEGWQRGNCSVPFPPTG 485

Qy 491 WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGT YGLNCAERCDCSHADGC 550
| | | | : | | | : | | | | | | | | : | | | | | | | | | | | | |
Db 486 WGFSCNASCQCAHEAVCS PQTGACTCTPGWHGAHCQLPCPKGFGE GCASRCDCHSDGC 545

Qy 551 HPTTGHCRCLPGW SGVHCD SVCAEGRWGPNC SLPCYCKNGASCSPDDGICECAPGFRGTT 610
| | | : | | | | | | | | | | | | | | | | : | : | | | | | | | |
Db 546 DPVHGRCQCQAGWMGARCHLSCEGLWGVNCSNTCTCKNGGTCLPENGNCVCAPGFRGPS 605

Qy 611 CQRICSPGFYGHRC SQTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA 670
| | | | | | | | | | | | | | | | : | |
Db 606 CQRSCQPGRYGKR-----CVP----- 621

Qy 671 GICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGE 730

Db 622 --CKCANHSFCHPSNGTCYCLAGWTGPDCSQPCPPGHWGENCAQTCQCHHGGTCHPQDGS 679
 Qy 731 CKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKCP 790
 Db 680 CICPLGWTGHHCLLEGCPGLTGFANCSQPCQCGPGEKC-----HPE----- 719
 Qy 791 SGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTALP 850
 Db 720 -----TGACVCPPGHSGAPCR-----IG-----IQEPFTVMP 746
 Qy 851 AD--SY-QIGAIAGIIILVLVLFLLALFIIYRHKQKGKESSMPAVTYTPAMRVVNADYT 907
 Db 747 TTPVAYNSLGAVIGIAVLGSLVVALVALFIGYRHWQKDKHHHLAVAYSSG-RLDGSEYV 805
 Qy 908 ISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLNKN-VN 966
 Db 806 MPDVPP-----SYSHYYSNPSYHTLSQCSPNPPPNK-----VPGPLFASLQNP 851
 Qy 967 PGKRGPVG-DCTGTLPADWKH-----GGYLNELGAFGLDRSYMKGSL-----KDLGK 1012
 Db 852 PG--GAQGHDNHTTLPADWKHRRPPPGPLDR-GSSRLDRSYSYSNPGPGPFYNKGLIS 908
 Qy 1013 NSEYNSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSP-----ARR 1057
 Db 909 EEELWASVASL-SSENPYATIRDLPGLPGGPRESSYMEMKGPPSGSPPRQPPQFWDQSRR 967
 Qy 1058 DSPYAEINNSTSANRNVYEVEPTVSVVQGVFSNNGRLSQDP-----YDLPKNSHIP 1108
 Db 968 RQPQPQRDSGT-----YE-QPSPL----IHDRDSVGSQPPLPPGLPPGHYDSPKNSHIP 1016
 Qy 1109 CHYDLLPVRDSSSSP-KQED 1127
 Db 1017 GHYDLPPVRHPPSPPLRRQD 1036

RESULT 11

US-09-796-753-114

; Sequence 114, Application US/09796753

; Publication No. US20030027998A1

; GENERAL INFORMATION:

; APPLICANT: McCarthy, Sean A.

; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF

; FILE REFERENCE: 7853-227-999

; CURRENT APPLICATION NUMBER: US/09/796,753

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 09/183,175

; PRIOR FILING DATE: 1998-10-30

; PRIOR APPLICATION NUMBER: 09/223,094

; PRIOR FILING DATE: 1998-12-30

; PRIOR APPLICATION NUMBER: 09/223,546

; PRIOR FILING DATE: 1998-12-30

; PRIOR APPLICATION NUMBER: 09/224,246

; PRIOR FILING DATE: 1998-12-30

; PRIOR APPLICATION NUMBER: 09/259,388

; PRIOR FILING DATE: 1999-02-26

; PRIOR APPLICATION NUMBER: 60/122,458

```

; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 114
; LENGTH: 1050
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-114

```

```

Query Match          37.2%; Score 2506.5; DB 10; Length 1050;
Best Local Similarity 40.5%; Pred. No. 7.7e-158;
Matches 490; Conservative 111; Mismatches 345; Indels 263; Gaps 30;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW----- 66
      ||| : | || || |||:: | :||: || : |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFWESFTTTTKESHSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      67 -----FKCTRHRVSYR-----TAY 80

```

Db	67	SPQTQRKLLASRDSFCMVCGAGVQWRDRSALQPQTGNALSMRPQPRVLSGAPSLASPGH	126
Qy	81	RHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGWGGTNCSSA--	138
Db	127	TVVVKTDHRQLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCVPGWRGDDCSSAPN	186
Qy	139	----CDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQR	194
Db	187	CLQPCTPGYYGPACQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQT-----	237
Qy	195	CQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGECS	254
Db	238	-----SGFFC-----PSTH-----PCQNGGVFQTPQGSCS	262
Qy	255	CPSGWMGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVG	314
Db	263	CPPGWMGTICSLPCPEGFHGPNCSEQECRCHNGGLCDRFTGQCRCAPGYTGDRCREECPPVG	322
Qy	315	TYGVLCAETCQCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLEN	374
Db	323	RFGQDCAETCDCAPDARCFPANGACLCEHGFTGDRCTDRLCPDGFGYGLSCQAPCTCDREH	382
Qy	375	THSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGF	434
Db	383	SLSCHPMNGECSCLPGWAGLHCNESCPODTHGPGCQEHCLCLHGGVCQATSGLCQCAPGY	442
Qy	435	KGIDCSTPCPLGTYGINCSSRCGCKNDAVCSVPDVGSTCKAGWHGVDCSIRCPSGTWGFG	494
Db	443	TGPHCASLCPDPTYGVNCSARCSCEAIACSPIDGECVCKEGWQRGNCSPVCPPGTWGFS	502
Qy	495	CNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTT	554
Db	503	CNASCQCAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASRCDCHSDGCDPVH	562
Qy	555	GHCRCPLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTTCQRI	614
Db	563	GRCQCQAGWMGARCHLSCPEGLWGVNCSNTCTCKNGGTCLPENGNCVCAPGFRGPPSCQRS	622
Qy	615	CSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICT	674
Db	623	CQPGRYGKR-----CVP-----CK	636
Qy	675	CTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCT	734
Db	637	CANHSFCHPSNGTCYCLAGWTGPDSCQPCPPGHWGGENCAQTCQCHHGGTCHPQDGSCICP	696
Qy	735	PGWTGLYCTQRCPLGFYKDCALICQCQNGADCDDHISGQCTCRTGFMGRHCEQKCPSTY	794
Db	697	LGWTGHHCLEGCPLGTFGANCSQPCQCGPGEKC-----HPE-----	732
Qy	795	YGGRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTALPAD--	852
Db	733	-----TGACVCPPGHSGAPCR-----IG-----IQEPFTVMPTTPV	763
Qy	853	SY-QIGAIAGIIILVLVLLFLLALFIIYRHKQKGKESSMPAVTYTPAMRVVNADYTISGT	911

Db 764 AYNLSGAVIGIAVLGSLVVALVALFIGYRHWQKGKEHHHLAVAYSSG-RLDGSEYVMPDV 822
 Qy 912 LPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKN-VNPGKR 970
 | : |||:|||||:|: :| | :|:| |
 Db 823 PP-----SYSHYYSNPSYHTLSQCSNP PPPPNK-----VPGPLFASLQNERPG-- 866
 Qy 971 GPVG-DCTGTL PADWKH-----GGYLNELGAFGLDRSYM GKSL-----KDLGKNSEY 1016
 | | | | | | | | | | : : : | | | | |
 Db 867 GAQGHDNHTTLPADWKHREPPPGPLDR-GSSRLDRSYSYSYNGPGPFYDKGLISEEEL 925
 Qy 1017 NSSNCSLSSSEN PYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYE 1076
 :| | | | | | | | :| | | :| | | :| | |
 Db 926 GASVASL-SSEN PYATIRDLP LPSLPGGPRESSYMEMKGPPSGSAPRQPPQFWDSQRRR--- 981
 Qy 1077 VEPTVSVVQGVFSNNGRL-----SQDP-----YDLPKNSHIPCHYDLLPVRDS 1119
 :| | : | | | | | | | | | | | | | | | | | |
 Db 982 -QPQPQRDSGTYEQPSPLIHDRDSVGSQPPLPPLPGHYDSPKNSHIPGHYDLPPVRHP 1040
 Qy 1120 SSSP-KQED 1127
 | | : : |
 Db 1041 PSPPLRRQD 1049

RESULT 12

US-10-052-648A-4

; Sequence 4, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Colman, Steven

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Gerlach, Valerie

; APPLICANT: Gunther, Erik

; APPLICANT: Kekuda, Ramesh

; APPLICANT: MacDougall, John R.

; APPLICANT: Mehraban, Fuad

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Shimkets, Richard

; APPLICANT: Smithson, Glennda

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Stone, David J.

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/284,549

```

; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-4

```

```

Query Match          36.7%; Score 2472.5; DB 15; Length 928;
Best Local Similarity 43.8%; Pred. No. 1.2e-155;
Matches 441; Conservative 100; Mismatches 313; Indels 153; Gaps 18;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSTCDILNW---FKCT 70
      ||| : | || || |||::|:|: || : | | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFEWESFTTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || || :||: ||| |||| | || || :|||::||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQRLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCQCVPGW 126

Qy     131 GGTNCSSACDGDHWGPHCTSRQCCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | ||| :|:| :||| :| : | | ||
Db     127 RGDDCSSECAPGMWGPQCQDKPCSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCPTPGYYGPA 186

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
      | |||| :|| || || || || || : | | | | |||||
Db     187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHSCQNGGVFQTPQ 245

Qy     251 GECSCPSGWM-----GTVCGQPCPEGRFGKNCSQECQCH 284
      | |||| ||| ||:| |||| | |||||:|
Db     246 GSCSCPPGWMVWRVGPVGMGCGSGENSVGGAKQGSKGITCSLPCPEGFHGPNCSSQECRCH 305

Qy     285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSGACLCEAG 344
      ||| || |||| |:||||:|:|:|||| :| |||| | :| :| ||||| |
Db     306 NGGLCDRFTGQCRCAPGYTGDRCREECPVGRFGQDCAETCDCAPDARCFPANGACLCEHG 365

Qy     345 FAGERCEARLCPEGLYGIKCDKRCPCHLENTHSCHPMSEGEACKPGWSGLYCNETCSPGF 404
      | |:| | ||||:| ||: | | :| :| ||||:|:| | ||:|:|:|
Db     366 FTGDRCTDRLCPDGFYGLSCQAPRTCDREHSLSCHPMNGECSCLPGWAGLHCNESCPQDT 425

Qy     405 YGEACQQICSCQNGADCDSVTGKCTCAPGFGKIDCSTPCPLGTYGINCSSRCGCKNDAVC 464

```

```

      :| ||: | | :| | : :| | |||: | | :| || |||:||||:| | :| |
Db      426 HGPCQEHCLCLHGGVCQATSGLCQCAPGYTGPHCASLCPDITYGVNCSARCSCEAIAC 485

Qy      465 SPVDGSCTCKAGWHGVDCSIRCPSTWGFSGCNLTCQCLNGGACNTLDGTCTCAPGWRGEK 524
      ||:|| | || || :||: || |||| || :|| : | :| ||| ||| |
Db      486 SPIDGECVCKEGWQRGNCSPVCPPTWGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAH 545

Qy      525 CELPCQDGTYGLNCAERCDCHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLP 584
      |:||| | :| || |||| :||| | | :| || | | | ||| |||
Db      546 CQLPCPKQGFGECCASRCDCHSDGCDPVHGRCCQAGWMGARCHLSCPEGLWGVNCSNT 605

Qy      585 CYCKNGASCSPDDGICECAPGFRGTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHITGL 644
      | |||| :| ||:| | ||||| :||| | || || |
Db      606 CTCKNGGTCLPENGNCVCAPGFRGPSCQRCQPGRYGKR----- 644

Qy      645 CDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCP 704
      |:|| | | | :| :| : :| | || | |||||
Db      645 --CVP-----CKCANHSFCHPSNGACYCLAGWTGPDCSQPCP 679

Qy      705 PAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNG 764
      | ||| || || ||:| | || | | |||| :| : |||| :| :|: ||| |
Db      680 PGHWGENCAQTCQCHHGGTCHPQDGSICPLGWTGHHCLLEGCPLGTFGANCSQPCQCGPG 739

Qy      765 ADCDHISGQCTCRTCRTGFMGRHCEQKCPSTYGYGCRQICDCLNNSTCDHITGTCYCSPGWK 824
      | | | | | | | |
Db      740 EKC-----HPE-----TGACVCPPGHS 756

Qy      825 GARCDQAGVIIVGNLNSLSTALPAD--SY-QIGAIAGIIILVLVVLFLALFIIYRH 881
      || | :| : | :| :| :||: || :| :|: |||| |||
Db      757 GAPCR-----IG-----IQEPFTVMPTTPVAYNSLGAVIGIAVLGSLVVALVALFIGYRH 806

Qy      882 KQKGKESMPAVTYTPAMRVVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPH 941
      || || || | :| : :|| : | : |||:|||||||:|: :|
Db      807 WQKDKEHHHLAVAYSSG-RLDGSEYVMPDVPP-----SYSHYYSNPSYHTLSQCSPNPP 859

Qy      942 VNNRDRMTVTKSKNNQLEVN LKN-VNPGKRGPGVG-DCTGTLPADWKH 986
      | :| || :|| | | | | |||||
Db      860 PPNK-----VPGPLFASLQNPFRPG--GAQGHDNHTTLPADWKH 896

```

RESULT 13

US-10-052-648A-6

; Sequence 6, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steven
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gunther, Erik
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacDougall, John R.
; APPLICANT: Mehraban, Fuad
; APPLICANT: Patturajan, Meera

```

; APPLICANT: Rothenberg, Mark
; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/284,549
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/303,229
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,605
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/269,098
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: 60/264,159
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/271,855
; PRIOR FILING DATE: 2001-02-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-648A-6

```

```

Query Match          36.7%; Score 2472.5; DB 15; Length 928;
Best Local Similarity 43.8%; Pred. No. 1.2e-155;
Matches 441; Conservative 100; Mismatches 313; Indels 153; Gaps 18;

```

```

Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT 70
      ||| : | || || |||::| :||: || : | | |
Db      9 LLLAVGLRLAGTLNPSDNTCSFEWESFTTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || || :::| || |||| || || :|||::||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQRLQCCHGFIYESRGFCVPLCAQECVHGRCVAPNQCCVPGW 126

Qy      131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | ||| :|:| :| | :| : | | |
Db      127 RGDDCSSECAPGMWGPQCDKPCSCGNSSCDPKSGVCSCPSGLQPPENCLQPCPTPGYYGPA 186

```


Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
| | | | : | | | | | | | | : | | | | | | | |

Db 187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHSCQNGGVFQTPQ 245

Qy 251 GECSCPSGWM-----GTVCGQPCPEGRFGKNCSEQCQCH 284
| | | | | | | | | | | | | | | | | |

Db 246 GSCSCPPGWMVWRVGPVGMGCGSGENSVGGAQGSKGITCSLPCPEGFHGPNCSEQECRH 305

Qy 285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAG 344
| | | | | | | | : | | | | : | | | | : | | | | : | | | |

Db 306 NGGLCDRFTGQCRCAPGYTGDRCREECPVGRFGQDCAETCDCAPDARCFPANGACLCEHG 365

Qy 345 FAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSEACKPGWSGLYCNETCSPGF 404
| | : | | | | : | | : | | | | : | | : | | | | : | | : | | | |

Db 366 FTGDRCTDRLCPDGFYGLSCQAPRTCDREHSLSCHPMNGECSCLPGWAGLHCNESCQDT 425

Qy 405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
: | | : | | : | | : | | | | : | | : | | | | : | | : | |

Db 426 HGPQCQERCLCLHGGVCQATSGLCQCAPGYTGPHCASLCPDITYGVNCSARCSCEAIAIC 485

Qy 465 SPVDGSCTCKAGWHGVDCSIRCPSTWGFSGCNLTQCCLNGGACNTLDGTCTCAPGWRGEK 524
| | : | | | | : | | : | | | | : | | : | | | |

Db 486 SPIDGECVCKEGWQRGNCSPVCPPTWGFSCNASQCAHEAVCSPQTGACTCTPGWHGAH 545

Qy 525 CELPCQDGTYGLNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDVCAEGRWGPNCSLP 584
| : | | | : | | | | : | | | | | | | | | | | | | |

Db 546 CQLPCPKGQFGEGCASRCDCHSDGCDPVHGRQCQAGWMGARCHLSCPEGLWGVNCSNT 605

Qy 585 CYCKNGASCSPDDGICECAPGFRGTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHITGL 644
| | | | : | | : | | | | : | | | | | | | | | |

Db 606 CTCKNGGTCLPENGNCVCAPGFRGPSCQQRSCQPGRYGKR----- 644

Qy 645 CDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCPGWIGSDCSQPCP 704
| : | | | | : | | : | | : | | | | | | | |

Db 645 --CVP-----CKCANHSFCHPSNGACYCLAGWTGPDCSQPCP 679

Qy 705 PAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNG 764
| | | | | | | | : | | | | | | | | : | | : | | | |

Db 680 PGHWGENCAQTCQCHHGGTCHPQDGSICPLGWTGHHCLGECPLGTFGANCSQPCQCGPG 739

Qy 765 ADCDHISGQCTCRTGFMGRHCEQKCPSTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWK 824
| | | | | | | | | | | | | | | | | |

Db 740 EKC-----HPE-----TGACVCPPGHS 756

Qy 825 GARCDQAGVIIIVGNLNSLSTALPAD--SY-QIGAIAGIIILVLVLLFLLALFIIYRH 881
| | | : | : | : | : | | : | | : | | : | | | | |

Db 757 GAPCR-----IG-----IQEPFTVMPPTTPVAYNSLGAVIGIAVLGSLVVALVALFIGYRH 806

Qy 882 KQKGKESSMPAVTYTPAMRVVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPH 941
| | | | | | : | : : | : | : | | : | | | | : | | : |

Db 807 WQKDKEHHHLAVAYSSG-RLDGSEYVMPDVPP-----SYSHYYSNPSYHTLSQCSNP 859

Qy 942 VNNRDRMTVTKSKNNQLFVNLKN-VNPGKRGPGV-DCTGTLPADWKH 986
| : | | : | | | | | | | | | |

Db 860 PPNK-----VPGPLFASLQNPFRPG--GAQGHDNHTLPADWKH 986

RESULT 14

US-10-052-648A-2

; Sequence 2, Application US/10052648A

; Publication No. US20040005558A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David

; APPLICANT: Burgess, Catherine

; APPLICANT: Casman, Stacie

; APPLICANT: Colman, Steven

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Gerlach, Valerie

; APPLICANT: Gunther, Erik

; APPLICANT: Kekuda, Ramesh

; APPLICANT: MacDougall, John R.

; APPLICANT: Mehraban, Fuad

; APPLICANT: Patturajan, Meera

; APPLICANT: Rothenberg, Mark

; APPLICANT: Shimkets, Richard

; APPLICANT: Smithson, Glennnda

; APPLICANT: Spytek, Kimberly A.

; APPLICANT: Stone, David J.

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Zerhusen, Bryan D.

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-250 (CURA-550)

; CURRENT APPLICATION NUMBER: US/10/052,648A

; CURRENT FILING DATE: 2002-12-09

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,920

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/284,549

; PRIOR FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: 60/303,229

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,605

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/269,098

; PRIOR FILING DATE: 2001-02-15

; PRIOR APPLICATION NUMBER: 60/264,159

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/265,517

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/271,855

; PRIOR FILING DATE: 2001-02-27

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 97

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 1020

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-052-648A-2

Query Match 36.4%; Score 2451.5; DB 15; Length 1020;
Best Local Similarity 40.6%; Pred. No. 3.3e-154;
Matches 481; Conservative 110; Mismatches 348; Indels 247; Gaps 28;

```
Qy      14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSTCDILNW---FKCT 70
      ||| : | || ||| ||| |||:: | :||: || : | | |
Db      9 LLLAVGLRLAGTLNPSDPNTCSFWESFTTTTKESHRSRPFSLLPSEPCE--RPWEGPHTCP 66

Qy      71 RHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
      : | ||| || ||| :||: ||| |||| ||| ||| :||| ||| ||| |||
Db      67 QPTVVYRTVYRQVVKTDHRQLQCCHGFYESRGFCVPLCAQECVHGRCVAPNQCCVPGW 126

Qy     131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNBITGACHCAAGFRGWRCEDRCEQGTYGND 190
      | :||| | ||| | | | : ||: | | : | : | | | | |
Db     127 RGDDCSSECAPGMWGPQCDKPCSCGNNSSCDPKSGVCSCPSGLQPPNCLQPCPTPGYYGPA 186

Qy     191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
      | |||| :|| || ||| || || ||| : | | | | | |||||
Db     187 CQFRCQC-HGAPCDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHSCQNGGVFQTPQ 245

Qy     251 GECSCPSGWM-----GTVCGQPCPEGRFGKNCSQECQCH 284
      | |||| ||| ||:| |||| | |||||:|
Db     246 GSCSCPPGWMVVRVGPVGMGCGSGENSVGGAQGSKGTICSLPCPEGFHGPNCSEQECRCH 305

Qy     285 NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSACLCEAG 344
      ||| || |||| |:||||:|:||||| :| |||| | :| :||| ||
Db     306 NGGLCDRFTGQCRCAPGYTGDRCREECPVGRFGQDCAETCDCAPDARCFPANGACLCEHG 365

Qy     345 FAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMSEGEACKPGWSGLYCNETCSPGF 404
      | :|| ||||:| ||: | | | :| ||||:| ||:| ||:| ||:|
Db     366 FTGDRCTDRLCPDGFYGLSCQAPCTCDREHSLSCHPMNGECSCLPGWAGLHCNESCPODT 425

Qy     405 YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC 464
      :| ||: | | :| | : :| | ||||: | | :| || ||||:| || :| |
Db     426 HGPQCQEHCLCLHGGVCQATSGLCQCAPGYTGPHCASLCPDPTYGVNCSARCSCENAIAC 485

Qy     465 SPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEK 524
      ||:| | || || :||: || |||| || :||: | : | ||| ||| |
Db     486 SPIDGECVCKEGWQRGNCVPCPPGTWGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAH 545

Qy     525 CELPCQDGTYGLNCAERCDCSHADGCHPTTGHCRCPLPGWSGVHCDVCAEGRWGPNCSLP 584
      |:|| | :| || |||| |:|| | | :| || | | || || |||
Db     546 CQLPCPKGQFGEGCASRCDCHSDGCDPVHGRCQCAQAGWMGARCHLSCEGLWGVNCSNT 605

Qy     585 CYCKNGASCSPDDGICECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPFCHHITGL 644
      | |||| :| ||:| | ||||| :||| | || || |
Db     606 CTCKNGGTCLPENGNCVCAPGFRGFPSCQQRSCQPGRYGKR----- 644

Qy     645 CDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCPGWIGSDCSQPCP 704
      | :|
Db     645 --CVP----- 647

Qy     705 PAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNG 764
      | | | :|| :| | | |||| |:||||| :| :||: ||| |
Db     648 -----CKCANHSFCHPSNGTCYCLAGWTGPDCSQRCPLGTFGANCSQPCQCGPG 696
```



```

; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 100
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Rauttus sp.
US-09-796-753-100

```

```

Query Match          28.3%; Score 1909; DB 10; Length 636;
Best Local Similarity 45.1%; Pred. No. 2e-118;
Matches 328; Conservative 77; Mismatches 212; Indels 110; Gaps 9;

```

```

Qy      260 MGTVCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVL 319
        || :| |||| | ||:||||:|||| || |||||:|||| |:|:|:||||| :|
Db      1  MGVICSLPCPEGFHGPNECTQECRCHNGGLCDRFTGQCHCAPGYIGDRCREECPVGRFGQD 60

Qy      320 CAETCQCVNGGKCYHVSAGACLCEAGFAGERCEARLCPGGLYGIKCDKRCPCHELENTHSCH 379
        |||| | | :|: :|||| || |:| ||||:| ||: | | | :| ||
Db      61 CAETCDCAPGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQDPCTCDPEHSLSCH 120

Qy      380 PMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDC 439
        || |||:|:||||:|:|:| :| ||: | | :| | :| | |||: | |
Db      121 PMHGECSQPGWAGLHCNESCQDTHGAGCQEHCLCLHGGVCLADSGLCRCAPGYTGPHC 180

```

Qy 440 STPCPLGTYGINCSSRCGCKNDVAVCSVDGSGCTCKAGWHGVDCSIRCPSGTWGFGCNLTC 499
: | | | | | | | | | | : | | | | | : | | | | | : |
Db 181 ANLCPPNTYGINCSSHCSCENAIACSPVDGTICKEGWQRGNCSVPCPPGTWGFSCNASC 240

Qy 500 QCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGHCRC 559
| | : | | : | | | | | | | : | | | | | : | | | | |
Db 241 QCAHEGVCSPTGTACTCTPGWRGVHCQLPCPKGQFGEGCASVDCDCHSDGCDPVHGHCR 300

Qy 560 LPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTTCQRICSPGF 619
| | | | | | | | | | | | : | : | | | | | : | | | | |
Db 301 QAGWMGTRCHLPCPEGFWGANCSNACTCKNGGTCVPENGNCVCAPGFRGPSCQRPCPPGR 360

Qy 620 YGHRCSTQCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNG 679
| | | | | | | | | | | | : | | | | | | | | | | |
Db 361 YGKR-----CVP-----CKCNNHS 374

Qy 680 TCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTG 739
: | : | | : | | | | | | | | | | | | | | | | | |
Db 375 SCHPSDGTCSCLAGWTGPDCESECPGHWGLKCSQPCQCHHGATCHPQDGSCVCI PGWTG 434

Qy 740 LYCTQRCPLGFYKGKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKCPSTYGYGCR 799
| : | | : | : | : | | | | | | | | | | | | | | |
Db 435 PNCSEGCPSRMFGVNCSQLCQCDPGEMC-----HPE----- 465

Qy 800 QICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTALPADSYQIGAI 859
| | | | | | | | | | | | : | : | : | : | : | : |
Db 466 -----TGACVCPGHS GAHCK-----VGSQESFTIMPTS-PVIHNSLGAV 504

Qy 860 AGIIILVLVVLFLALFIIYRHKQKGKESMPAVTYTPAMRVVNADYTISGTLPHSNGGN 919
| | : | : | : | | | | | | | | | | | | : | : | : | |
Db 505 IGI AVLGLVVALVALFIGYRHWQKGKEHEHLAVAYSTG-RLDGS DYVMPDVSP----- 557

Qy 920 ANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKNVNPGRGPVGDCTGT 979
: | | : | | | | | : | : | : | | | | | | : | | : | | | |
Db 558 SYSHYYSNPSYHTLSQCSNP P P P P N-----KIPGSQLFVSSQASERPNNRHGRDNHAT 610

Qy 980 LPADWKH 986
| | | | | | | |
Db 611 LPADWKH 617

Search completed: March 26, 2004, 16:21:15
Job time : 59.8389 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 16:04:46 ; Search time 55.4809 Seconds
(without alignments)
6483.148 Million cell updates/sec

Title: US-10-092-390-2
Perfect score: 6744
Sequence: 1 MVISLNSCLSFICLLCHWI.....SSPKQEDSGGSSSNSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriaphage:*
17: sp_archaeophages:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							

1	6744	100.0	1140	4	Q96KG7	Q96kg7 homo sapien
2	3769	55.9	969	4	Q96KG6	Q96kg6 homo sapien
3	3509.5	52.0	947	11	Q8BKK7	Q8bkk7 mus musculu
4	3468	51.4	567	4	Q8WUL3	Q8wul3 homo sapien
5	3311.5	49.1	921	11	Q80T91	Q80t91 mus musculu
6	2668	39.6	1034	11	Q8VHL7	Q8vhl7 mus musculu
7	2667	39.5	1034	11	Q8VIK5	Q8vik5 mus musculu
8	2529	37.5	1004	11	Q8CGA7	Q8cga7 mus musculu
9	2292.5	34.0	747	11	Q8VHF4	Q8vhf4 mus musculu
10	2135.5	31.7	626	4	Q8ND91	Q8nd91 homo sapien
11	1958	29.0	1574	11	O88281	O88281 rattus norv
12	1931	28.6	1664	5	Q9TVQ2	Q9tvq2 caenorhabdi
13	1832.5	27.2	881	5	Q9W0A0	Q9w0a0 drosophila
14	1823.5	27.0	1070	5	Q8T3A7	Q8t3a7 caenorhabdi
15	1805.5	26.8	1111	5	Q9XWD6	Q9xwd6 caenorhabdi
16	1804.5	26.8	1246	4	O75095	O75095 homo sapien
17	1804	26.7	1045	5	Q8T3A6	Q8t3a6 caenorhabdi
18	1401	20.8	546	11	Q80V70	Q80v70 mus musculu
19	1341	19.9	299	11	Q8BX64	Q8bx64 mus musculu
20	1309	19.4	220	11	Q63404	Q63404 rattus norv
21	1041	15.4	2516	11	Q7TQ52	Q7tq52 mus musculu
22	1041	15.4	2526	11	Q7TQ51	Q7tq51 mus musculu
23	1041	15.4	2531	11	Q8K428	Q8k428 mus musculu
24	1041	15.4	2531	11	Q7TQ50	Q7tq50 mus musculu
25	1033	15.3	2428	5	Q8I6X6	Q8i6x6 boophilus m
26	1024.5	15.2	2447	13	O13149	O13149 fugu rubrip
27	1014	15.0	2468	13	Q800E4	Q800e4 brachydanio
28	1006	14.9	2653	5	Q25253	Q25253 lucilia cup
29	988.5	14.7	2524	5	Q9GPA5	Q9gpa5 branchiosto
30	981	14.5	2528	13	Q8AXP0	Q8axp0 cynops pyrr
31	964.5	14.3	2531	5	O16004	O16004 lytechinus
32	920.5	13.6	4006	11	O35452	O35452 mus musculu
33	916.5	13.6	491	4	Q8TEK2	Q8tek2 homo sapien
34	916.5	13.6	4288	4	Q9NPK9	Q9npk9 homo sapien
35	915.5	13.6	2352	5	O61240	O61240 halocynthia
36	903.5	13.4	4114	11	O54796	O54796 mus musculu
37	881.5	13.1	4135	6	O18977	O18977 bos taurus
38	874	13.0	1918	5	Q86AS3	Q86as3 dictyosteli
39	870.5	12.9	569	4	Q8NHD4	Q8nhd4 homo sapien
40	863	12.8	713	5	Q962W9	Q962w9 podocoryne
41	839	12.4	320	4	Q8N780	Q8n780 homo sapien
42	832.5	12.3	752	13	O42374	O42374 brachydanio
43	800	11.9	866	4	Q8IXF3	Q8ixf3 homo sapien
44	790.5	11.7	1214	13	Q90YD2	Q90yd2 xenopus lae
45	782	11.6	1254	13	Q9YHU2	Q9yhu2 brachydanio

ALIGNMENTS

RESULT 1

Q96KG7

ID Q96KG7 PRELIMINARY; PRT; 1140 AA.

AC Q96KG7;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE MEGF10 protein (Hypothetical protein KIAA1780).
 GN MEGF10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Hippocampus;
 RX MEDLINE=21245130; PubMed=11347906;
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large Proteins in vitro.";
 RL DNA Res. 8:85-95(2001).
 DR EMBL; AB058676; BAB47409.1; -.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 10.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 6.
 DR PROSITE; PS00022; EGF_1; 17.
 DR PROSITE; PS01186; EGF_2; 17.
 KW Hypothetical protein; EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 1140 AA; 122204 MW; 45B2FA239423895A CRC64;

Query Match 100.0%; Score 6744; DB 4; Length 1140;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	1	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Qy	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Db	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCED	180
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPPGKHGPPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPPGKHGPPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Db	301	GYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLY	360
Qy	361	GIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420

Db	361	GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD	420
Qy	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV	480
Db	421	CDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCTCKAGWHGV	480
Qy	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Db	481	DCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAE	540
Qy	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGIC	600
Db	541	RCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGIC	600
Qy	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVC	660
Db	601	ECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVC	660
Qy	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Db	661	PSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHN	720
Qy	721	GAFC SAYDGECKCTPGWTGLYCTQRCPLGFYGKD CALICQCQNGADCDHISGQCTC RTGF	780
Db	721	GAFC SAYDGECKCTPGWTGLYCTQRCPLGFYGKD CALICQCQNGADCDHISGQCTC RTGF	780
Qy	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGVII VGNLN	840
Db	781	MGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGVII VGNLN	840
Qy	841	SLSRTSTALPADSYQIGAIAGIIILVVLVFL LALFIIYRHKQKGKESSMPAVTYTPAMR	900
Db	841	SLSRTSTALPADSYQIGAIAGIIILVVLVFL LALFIIYRHKQKGKESSMPAVTYTPAMR	900
Qy	901	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVT KSKNNQLFV	960
Db	901	VVNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVT KSKNNQLFV	960
Qy	961	NLKNVNPGRGPVGDCTGTL PADWKHGGYLNELGAFGLDRSYM GKSLKDLGKNSEYNSSN	1020
Db	961	NLKNVNPGRGPVGDCTGTL PADWKHGGYLNELGAFGLDRSYM GKSLKDLGKNSEYNSSN	1020
Qy	1021	CSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTS ANRNVYEVEPT	1080
Db	1021	CSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINNSTS ANRNVYEVEPT	1080
Qy	1081	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1140
Db	1081	VSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSPKQEDSGGSSSNSSSSSE	1140

RESULT 2

Q96KG6

ID Q96KG6 PRELIMINARY; PRT; 969 AA.

AC Q96KG6;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

```

DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      MEGF11 protein (Hypothetical protein KIAA1781).
GN      MEGF11.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=21245130; PubMed=11347906;
RA      Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT      "Prediction of the coding sequences of unidentified human genes. XX.
RT      The complete sequences of 100 new cDNA clones from brain which code
RT      for large Proteins in vitro.";
RL      DNA Res. 8:85-95(2001).
DR      EMBL; AB058677; BAB47410.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR005805; Rieske.
DR      Pfam; PF00008; EGF; 12.
DR      PRINTS; PR00011; EGFLAMININ.
DR      SMART; SM00180; EGF_Lam; 8.
DR      PROSITE; PS00022; EGF_1; 17.
DR      PROSITE; PS01186; EGF_2; 17.
DR      PROSITE; PS00200; RIESKE_2; 1.
KW      Hypothetical protein; EGF-like domain; Laminin EGF-like domain.
SQ      SEQUENCE 969 AA; 101600 MW; 56DD2FFE139C8209 CRC64;

```

Query Match 55.9%; Score 3769; DB 4; Length 969;
Best Local Similarity 58.6%; Pred. No. 2.9e-292;
Matches 600; Conservative 126; Mismatches 208; Indels 90; Gaps 4;

Qy	109	CADKCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPIITGACH	168
Db	28	CTEECVHGRCSVSPDTCHEPGWGGPDCSSGCDSDHWGPHCSNRCQCQNGALCNPIITGACV	87
Qy	169	CAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPPG	228
Db	88	CAAGFRGWRCEELCAPGTHGKGCLPCQCRHGASCDPRAGECLCAPGYTGVYCEELCPPG	147
Qy	229	KHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGPCPEGRFGKNCSQECQCHNGGT	288
Db	148	SHGAHCELRCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSDCPCHHGGQ	207
Qy	289	CDAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNNGGKCYHVSAGACLCEAGFAGE	348
Db	208	CDHVTGQCHCTAGYMGDRCQECPFGSFGFQCSQRCDCHNNGGQCSPTTGACECEPGYKGP	267
Qy	349	RCEARLCPEGLYGIKCDKRCPCHLNTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEA	408

Db 268 RCQERLCPEGLHGPCTLPCCDADNTISCHPVTGACTCQPGWSGHHHCNESCVPVGYGDG 327

Qy 409 CQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSFVD 468
 || |:||||| |:|| | |||| | | | | ||||

Db 328 CQLPCTCQNGADCHSITGGCTCAPGFMGEVCAVSCAAGTYGPNCSICSCNNGGTCSFVD 387

Qy 469 GSCTCKAGWHGVDCSIRCPSTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELP 528
 ||||| || |:|::: ||||| || :| | | | |: |:|:| || |: |||

Db 388 GSCTCKEGWQGLDCTLPCTSGTWGLNCNESCTCANGAACSPIDGSCSCTPGWLGDTCCLP 447

Qy 529 CQDGTYGLNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDSDVCAEGRWGPNCSLPCYCK 588
 | |||:||||:| ||||| | || | | |:|: || | | |||||: | |:

Db 448 CPDGTFTGLNCEHCDSCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNCVSVCSCE 507

Qy 589 NGASCSPDDGICECAPGFRGTTCQRICSPGFYGHRCSTCPQCVHSSGPFCHHITGLCDCL 648
 || ||||:|| ||||| |||| | |||| |:| || |||| | ||||:|:|:

Db 508 NGGSCSPEDGSCECAPGFRGPLCQRICPPGFYGHGCAQPCPLCVHSSRPCHHISGICECL 567

Qy 649 PGFTGALCNEVCPSTGRFGKNCAGICTCTNNGTCNPIDRSCQCPGWIGSDCSQPCPPAHW 708
 |||:||||:| | |:|:| |:| | ||||:| | |||:| | | | || |

Db 568 PGFSGALCNQVCAGGYFGQDCAQLCSCANNGTCSPIDGSCQCFPGWIGKDCSQACPPGFW 627

Qy 709 GPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKDCALICQCCQNGADCD 768
 || | |:|||| | | | | |||||:|||| |:||| :|||| |

Db 628 GPACFHACSCHNGASCSAEDGACHCTPGWTGLFCTQRCPAAFFGKDCGRVCQCCQNGASCD 687

Qy 769 HISGQCTCRTGFMGRHCEQKCPSTGYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARC 828
 |||:||||| |:|||:| |:||||:|:|:| |||||:|||||:| |

Db 688 HISGKCTCRTGFTGQHCEQRCAPGTFTGYGCQQLCECMNNSTCDHVTGTCTCYCSPGFKGIRC 747

Qy 829 DQAGVIIIVGNLNSLSRTSTALPADSYQIGAIAGIIILVLVFLFLLALFIIYRHKQKGKES 888
 ||| ::: || :: | || |: : |:| |:|: : : || | |: :| |

Db 748 DQA-ALMMEELNPYTKISPALGAERHVSVAVTGIMLLFFIVVLLGLFAWHRRRQKEKGR 806

Qy 889 SM-PAVTTYTPAMRVVNADYTIISGTLPHSNNGGNANSHYFTNPSYHTLTQCATSPHVNNRDR 947
 : | |:|||||: : ||::|

Db 807 DLAPRVSYTPAMRMTSTDYSLS----- 828

Qy 948 MTVTKSKNNQLFVNLKVNPNPKRGFVGDCGTGLPADWKHGGYLNELGAFGLDRS----YM 1003
 || |:| |

Db 829 -----GACGMDDRRQNTYIM 842

Qy 1004 GKSLKDLGKNSEYNSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAE 1063
 | || | |:| |||:|||||||:| | | ||||| |||:

Db 843 DKGFKDYMKESVCSSTCSLNSSENPYATIKDPPILTCKLPESYVEMKSPVHMGSPYTD 902

Qy 1064 INNSTSANRNVYEVEPTVSVVQGVFSNNGRLSQDPYDLPKNSHIPCHYDLLPVRDSSSSP 1123
 : : ::|:|:||||||| :| | |:|:| ||||| | |:

Db 903 VPSLSTSNKNIYEVEPTVSVVQEGCGHNSSYIQNAYDLPRNSHIPGHYDLLPVRQSPANG 962

Qy 1124 KQED 1127
 :|

Db 963 PSQD 966

RESULT 3
 Q8BKK7

ID Q8BKK7 PRELIMINARY; PRT; 947 AA.
AC Q8BKK7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MEGF11 protein.
GN 2410080H04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK051642; BAC34702.1; -.
DR MGD; MGI:1920951; 2410080H04Rik.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR005805; Rieske.
DR Pfam; PF00008; EGF; 11.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00180; EGF_Lam; 15.
DR PROSITE; PS00022; EGF_1; 15.
DR PROSITE; PS01186; EGF_2; 15.
DR PROSITE; PS00200; RIESKE_2; 1.
SQ SEQUENCE 947 AA; 100661 MW; 0C209B11DFEE8314 CRC64;

Query Match 52.0%; Score 3509.5; DB 11; Length 947;
Best Local Similarity 56.1%; Pred. No. 1.5e-271;
Matches 577; Conservative 122; Mismatches 217; Indels 113; Gaps 13;

Qy 15 LLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHFDQIYYTSCTDILNWFKCTRHRV 74
|| : | || ||||| : ||||| ||||| | ||||| :
Db 8 LLVFLQLAALALNPEDPNVCSHWESYAVTVQESYAHFPDQIYYTRCADILNWFKCTRHRV 67
Qy 75 SYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGWGGTN 134
||:|||| | :||||:|||||:|:|: |:
Db 68 SYKTAYRRGLRTMYRRRSQCCPGYYENGDFCI----- 99
Qy 135 CSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQR 194
|| :|||||:||||:||||| || |||||: | ||: |
Db 100 ---RCDSEHWGPHCSNRCQCGNGALCNPITGACVCAPGFRGWRCEELCAPGTHGKGCQLL 156
Qy 195 CQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVTGEC 254

Db 157 CQCHHGASCDPRTGECLCAPGYTGVYCEELCPPGSHGAHCELRCPQNGGTCHHITGECA 216
 QY 255 CP SGWMTVC GQPCPEGRFGKNCSQECQCHNGGTCDAA TGQCHCSPGYTGERCQDECPVG 314
 Db 217 CPPGWTGAVCAQPCPPGTFGQNCSDCPCHHGGQCDHVTGQCHCTAGYMGDRCQEECPFG 276
 QY 315 TYGVLCAETCQCVNNGGKCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLEN 374
 Db 277 TFGFLCSQRCDCHNGGQCSPATGACECEPGYKGPSCQERLCPEGLHGP GCTLPCPCD TEN 336
 QY 375 THSCHPMMSGECACKPGW SGLYCNETCSPGFYGEACQQICSCQNGADCD SVTGKCTCAPGF 434
 Db 337 TISCHPVTGACTCQPGW SGHYCNESCPAGYYNGCQLPCTCQNGADCHSITGSCTCAPGF 396
 QY 435 KGIDCSTPCPLGTYGINCSSRCGCKNDAVCS PVDG SCTCKAGWHGVDCSIRCPSGTWGFG 494
 Db 397 MGEVCAVPCAAGTYGPNCSSVCSCSNGGTCS PVDG SCTCREGWQGLDCSLPCPSGTWGLN 456
 QY 495 CNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGT YGLNCAERCD CSHADGCHPTT 554
 Db 457 CNETCICANGAACSPFDGSCACTPGWLGDSCELPCPDGT FGLNCSEHCD CSHADGCDPVT 516
 QY 555 GHCRCLPGW SGVHCD SVCAEGRWGPNC SLPCYCKNGASCSPDDGICECAPGFRGTT CQRI 614
 Db 517 GHCCCLAGWTGIRCDSTCPPGRWGPNC SVSCSENGGSCSPEDGSC ECAPGFRGPLCQRI 576
 QY 615 CSPGFYGHRC SQTCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICT 674
 Db 577 CPPGFYGHGCAQPCPLCVHSRGPCHHISGICECLPGFSGALCNQVCAGGHFGQDCAQLCS 636
 QY 675 CTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCT 734
 Db 637 CANNGTCSPIDGSCQCFPGWIGKDCSQGCPSA----- 668
 QY 735 PGWTGLYCTQRCPLGFY GKDCALICQ CQNGADCDHISGQCTC RTGFMGRHCEQKCPSGTY 794
 Db 669 -----FFGKDCGHICQ CQNGASCDHITGKCTC RTGFSGRHCEQRCAPGTF 713
 QY 795 GYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGV IIVGNLNSLSRTSTALPADSY 854
 Db 714 GYG CQQ LCECMN NATCDHVTGTCYCSPGFKGIRCDQA-ALMMDELNPYTKISPALGAERH 772
 QY 855 QIGAIAGI IILVLVVLFL LALFIIYRHKQKGKES SM-PAV TYTPAMRVVNADY TISGTL P 913
 Db 773 SVGAVTGIVLLLFLVVL LGLFAWRRRRQKEKGRDLAPRVSYTPAMRMTSTDYSLSDL-- 830
 QY 914 HSNNGNANSHYFTNPSYHTLTQC---ATSPHVNNRDRMTVTKSKNNQLFVN LKNVNP GKR 970
 Db 831 --SQSSSHAQCFSNASYHTLA-CGGPAT S-QASTLDRNSPTKLSNKS L-----DR 876
 QY 971 GPVGDC TGTLPADWKHGGYLNELGAF----GLDRSYM GK----SLKDLGKNSE-YNSSNC 1021
 Db 877 DTAG-----WTPYSYVNVLD SHFQISALEARYPPEDFYIELRHLSRHAEPHSPGTC 927
 QY 1022 SLSSSEN PY 1030
 : : | |

RESULT 4

Q8WUL3

ID Q8WUL3 PRELIMINARY; PRT; 567 AA.
 AC Q8WUL3;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to MEGF10 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC020198; AAH20198.1; -.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 7.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 4.
 DR PROSITE; PS00022; EGF_1; 10.
 DR PROSITE; PS01186; EGF_2; 10.
 KW EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 567 AA; 60797 MW; CF2FB8CDEB7CF627 CRC64;

Query Match 51.4%; Score 3468; DB 4; Length 567;
 Best Local Similarity 99.8%; Pred. No. 1.6e-268;
 Matches 565; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Db	1	MVISLNSCLSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSC	60
Qy	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Db	61	TDILNWFKCTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIA	120
Qy	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE	180
Db	121	PNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCE	180
Qy	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGHPQCEQRCPC	240
Db	181	RCEQGTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGHPQCEQRCPC	240
Qy	241	QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Db	241	QNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSP	300
Qy	301	GYTGERCQDECPVGTYGVLCAETCQCVNGGKCYHVSAGACLCEAGFAGERCEARLCPEG	360

```

Db      301  |||||||||||||||||||||||||||||||||||||||||||||||||||||||| 360
          GYTGERCQDECPVGTGYGLCAETCQCVNGGKCYHVSAGACLCEAGFAGERCEARLCPEGLY
Qy      361  GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361  GIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGAD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      421  CDSVTGKCTCAPGFKGIDCSTPCPLGTGYGINCSSRCGCKNDAVCSPVDGSCCTCKAGWHGV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421  CDSVTGKCTCAPGFKGIDCSTPCPLGTGYGINCSSRCGCKNDAVCSPVDGSCCTCKAGWHGV 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      481  DCSIRCPSTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAE 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481  DCSIRCPSTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYGLNCAE 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      541  RCDCSHADGCHPTTGHCRLPGWSGV 566
          ||||||||||||||||||||||||||||:
Db      541  RCDCSHADGCHPTTGHCRLPGWSGL 566

```

RESULT 5

Q80T91

```

ID   Q80T91          PRELIMINARY;          PRT;    921 AA.
AC   Q80T91;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MKIAA1781 protein (Fragment).
GN   MKIAA1781.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
DR   EMBL; AK122555; BAC65837.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR005805; Rieske.
DR   Pfam; PF00008; EGF; 10.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 14.

```



```

DR SMART; SM00180; EGF_Lam; 14.
DR PROSITE; PS00022; EGF_1; 14.
DR PROSITE; PS01186; EGF_2; 14.
DR PROSITE; PS00200; RIESKE_2; 1.
FT NON_TER      1      1
SQ SEQUENCE      921 AA;  97316 MW;  60A34D9513A600F7 CRC64;

```

Query Match 49.1%; Score 3311.5; DB 11; Length 921;
Best Local Similarity 59.6%; Pred. No. 9.6e-256;
Matches 552; Conservative 110; Mismatches 233; Indels 31; Gaps 10;

Qy	225	CPPGKHGPGQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSEQECQH	284
Db	1	CPPGSHGAHCELRCPCQNGGTCHHITGECACPPGWTGAVCAQPCPPGTFGQNCSEQDCPCH	60
Qy	285	NGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCACETCQCVNGGKCYHVSAGACLCEAG	344
Db	61	HGGQCDHVTGQCHCTAGYMGDRQEECPFGTFGFLCSQRCDCHNGGQCSPATGACECEPG	120
Qy	345	FAGERCEARLCPGEGLYGIKCDKRCPCCHLENTHSCHPMSGECACKPGWSGLYCNETCSPGF	404
Db	121	YKGPSQCERLCPEGLHGPCTLPCCDTEINTISCHPVTGACTCQPGWSGHYCNESCPAGY	180
Qy	405	YGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVC	464
Db	181	YGNGCQLPCTCQNGADCHSITGSCTCAPGFMGEVCAVPCAAGTYGPNCSSVSCSNGGTC	240
Qy	465	SPVDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEK	524
Db	241	SPVDGSCTCREGWQGLDCSLPCPSGTWGLNCNETCICANGAACSPFDGSCACTPGWLGDS	300
Qy	525	CELPCQDGTYGLNCAERCDCSHADGCHPTTGHCRCPLPGWSGVHCDSVCAEGRWGPNCSLP	584
Db	301	CELPCPDGTGFLNCSEHCDCSHADGCDPVTGHCCCLAGWTGIRCDSTCPPGRWGPNCVS	360
Qy	585	CYCKNGASCSPDDGICECAPGFRGTTCQRICSPGFYGHRCSTQCPQCVHSSGPCHHITGL	644
Db	361	CSCENGGSCSPEDGSCECAPGFRGPLCQRICPPGFYGHGCAQPCPLCVHSRGPCHHISGI	420
Qy	645	CDCLPGFTGALCNEVCPSGREGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCP	704
Db	421	CECLPGFSGALCNQVCAGGHFGQDCAQLCSCANNGTCSPIDGSCQCFPGWIGKDCSQACP	480
Qy	705	PAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCNQ	764
Db	481	SGFWSACFHTCSCHNGASCSAEDGACHCTPGWTGLFCTQRCPSAFFGKDCGHICQCNQ	540
Qy	765	ADCDHISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWK	824
Db	541	ASCDHITGKCTCRTGFSGRHCEQRCAPGTFGYGCQQLCECMNNATCDHVTGTCTCYCSPGFK	600
Qy	825	GARCDQAGVIVGNLNSLSRTSTALPADSYQIGAIAGIILVLVVLFLLLALEFIYRHKQK	884
Db	601	GIRCDQA-ALMMDELNPYTKISPALGAERHSVGAVTGIVLLLEFLVVVLLGLFAWRRRRQK	659
Qy	885	GKESSM-PAVITYTPAMRVVNADYTIISGTLPHSNGGNANSHYFTNPSYHTLTQC--ATSP	940

Db 660 EKGRDLAPRVSYTPAMRMTSTDYSLSDL-----SQSSSHAQCFSNASYHTLA-CGGPATS- 713

Qy 941 HVNNRDRMTVTKSKNNQLFVNLIKVNPGKRGPVGDC---TGTLPADWKHGGYLNEL----- 993
: || : || | | : | | | | : : ||

Db 714 QASTLDRNSPTKLSNKS LDRDTAGWTPYSYVNVLD SHFQISALEARYPPEDFYIELRHLS 773

Qy 994 -----GAFGLDRS-----YMGKSLKDLGKNSEYNSSNCSLSSSENPYATIKDPPVLIP 1041
| | :|| | | | | :|| |||:|||||:|

Db 774 RHAEPHSPGTCGMDRRQNTYIMDKGFKDYMKESVCSSTCSLNSSSENPYATIKDPPILTC 833

Qy 1042 KSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPTVSVVQGVFSNNGRLSQDPYDL 1101
| | ||||| ||| :: : ::||:||||| :| |:|||

Db 834 KLPESYVEMKSPVHLGSPYTDVPSLSTSNKNIYEVEPTVSVVQEGRGHNSSYIQNPYDL 893

Qy 1102 PKNSHIPCHYDLLPVRDS-SSSPKQE 1126
||||| ||||| | : | ||

Db 894 PKNSHIPGHYDLLPVRQSPAHPGFQE 919

RESULT 6

Q8VHL7

ID Q8VHL7 PRELIMINARY; PRT; 1034 AA.

AC Q8VHL7;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Jedi protein.

GN 3110045G13RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL; TISSUE=Testis;

RA Krivtsov A.V., Zinovyeva M.V., Hendrikx J., Visser J.W.M.,

RA Belyavsky A.V.;

RT "Jedi is a novel DSL and EGF-like repeat motif-containing protein

RT expressed on non-differentiated hematopoietic cells.";

RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF444274; AAL38571.1; -.

DR MGD; MGI:1920432; 3110045G13Rik.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR009030; Grow_fac_recep.

DR InterPro; IPR002049; Laminin_EGF.

DR Pfam; PF00008; EGF; 6.

DR PRINTS; PR00011; EGFLAMININ.

DR SMART; SM00180; EGF_Lam; 4.

DR PROSITE; PS00022; EGF_1; 13.

DR PROSITE; PS01186; EGF_2; 12.

KW EGF-like domain; Laminin EGF-like domain.

SQ SEQUENCE 1034 AA; 110540 MW; 5514E5166AE01111 CRC64;

Query Match 39.6%; Score 2668; DB 11; Length 1034;

Best Local Similarity 42.7%; Pred. No. 2.9e-204;

Matches 493; Conservative 110; Mismatches 379; Indels 172; Gaps 16;

Qy 14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSTCDILNW---FKCT 70
 ||| : || ||||: |||: | :||: || : || |
 Db 7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKEHLRPFSLPAESCH--RPWEDPHTCA 64

 Qy 71 RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
 : | ||| || | : ||| :||| ||| || :|||:||| ||| |||
 Db 65 QPTVVYRTVYRQVVKMDSRPRLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW 124

 Qy 131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
 | :||| | ||| | | | : |:| :||| | :| : | | |
 Db 125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDBPKSGACFCPSGLQPPNCLQPCPAGHYGPA 184

 Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
 | ||| ||:|| | |||| | | | | | : |||||
 Db 185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGDGFFCPRTYPCQNGGVFQGSQ 243

 Qy 251 GECSCPSGWMGTVCQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
 | |||| |||| :| |||| | ||:||:|||| || |||||:||| |:||:|
 Db 244 GSCSCPPGWMGVICSLPCPEGFHGPNTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

 Qy 311 CPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
 |||| :| |||| | | :| : |||| | ||:| ||||:| ||: | : | |
 Db 304 CPVGRFGQDCAETCDCAPGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

 Qy 371 HLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
 |:: |||| |||:|:|:|:|:|:| :| ||: | | :| | :| | |
 Db 364 DPEHSLSCHPMHGECSQCPGWAGLHCNESCQDTHGPGCQEHCLCLHGGLCLADSGLCRC 423

 Qy 431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDVCSVVDGSCCKAGWHGVDCSIRCPSTG 490
 |||: | |: || ||||| ||| |:| |||:|:| || || :||: || ||
 Db 424 APGYTGPHCANLCPDPTYGINCSSRCSCEAIACSPIDGTICKEGWQRGNCSVPCPLGT 483

 Qy 491 WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGILNCAERCDCSHADGC 550
 ||| || :|| : | |: | ||| || | |:|| | :| || ||| |:|||
 Db 484 WGFNCNASCQCAHDGVCSPTGTACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 543

 Qy 551 HPTTGHCRCCLPGWSGVHCDVCAEGRWGPNCSLPCYCKNGASCSRDDGICECAPGFRGTT 610
 | | ||| || | | || || || | |||| :| :| | ||||| :
 Db 544 DPVHGQCRCQAGWMGTRCHLPCPEGFWGANCSNTCTCKNGGTVCSENGNCVAPGFRGPS 603

 Qy 611 CQRICSPGFYGHRCSTCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA 670
 ||| | || || || |
 Db 604 CQRPCPPGRYGRKRCVQ----- 619

 Qy 671 GICTCTNN-GTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDG 729
 | | || :|:| | :| | || | |||: ||| || | |||:| | ||
 Db 620 --CKCNNHSSCHPSDGTCSCLAGWTGPDCEACPPGHWGLKCSQLCQCHHGGTCHPQDG 677

 Qy 730 ECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKC 789
 | ||||| | : || :| :|: ||| |
 Db 678 SCICTPGWTGPNCLEGCPPRMFGVNCSQLCQCDLG----- 712

 Qy 790 PSGTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGARCDQAGVIVGNLNSLSRTSTAL 849
 | ||| | || | | :|: | : | :
 Db 713 -----EMCHPQTGACVCP PGHSGADCK-----MGSQESFTIMPTS- 747

Qy 850 PADSYQIGAIAGIIILVLVLFLLALFIIYRHKQKGKESSMPAVTYTPAMRVVNADYTIS 909
 | :||: || :| :|: |:|||| || ||||| || :| :| :|| :
 Db 748 PVTHNSLGAVIGIAVLGTLVVALIALFIGYRQWQKGKEHEHLAVAYSTG-RLDGS DYVMP 806

Qy 910 GTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKNVNP GK 969
 | : |||: |||||: ||: :| | : |||: :
 Db 807 DVSP-----SYSHYYSNPSYHTLSQCSNP PPPN-----KVPGSQLFVSSQAPERPS 853

Qy 970 RGPVGDCGTLPADWKHGGYLNELGAFGLDRSY-----MGKS 1006
 | : ||||| || :| || ||||| :| |
 Db 854 RAHGRENHVTL PADWKHRRPHERGASHLDRSYSCSYSHRNGPGPFCHKGPIS E E GLGAS 913

Qy 1007 LKDLGKNSEYNSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINN 1066
 : | ||||| ||: | | : | ||||| | | : :
 Db 914 VMSL-----SSENPYATIRDLPSLPGEPR ESGYVEMKGPPSVSPPRQSLH- 958

Qy 1067 STSANRNVYEVEP-----TVSVVQGVFSNNGRLSQDP-----YDLPKNSHIPCHY 1111
 :| :||: | : | | | || ||||| ||
 Db 959 --LRDRQQRQLQPQRDSGTYEQPSPLSHNEESLGSTPPLPPGLPPGQYDSPKNSHIPGHY 1016

Qy 1112 DLLPVRDSSSSPKQ 1125
 || ||| | | :
 Db 1017 DLPPVRHPPSPPSR 1030

RESULT 7

Q8VIK5

ID Q8VIK5 PRELIMINARY; PRT; 1034 AA.
 AC Q8VIK5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MEGF12.
 GN 3110045G13RIK OR MEGF12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RA Ivanova N.B., Lemischka I.R.;
 RT "The global gene expression profiling of the hematopoietic stem
 RT cell.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Eye;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL; AF440279; AAL33583.1; -.
 DR EMBL; AK053551; BAC35426.1; -.
 DR MGD; MGI:1920432; 3110045G13Rik.

DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 6.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 4.
 DR PROSITE; PS00022; EGF_1; 13.
 DR PROSITE; PS01186; EGF_2; 12.
 KW EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 1034 AA; 110580 MW; 714E5016848E4E4C CRC64;

Query Match 39.5%; Score 2667; DB 11; Length 1034;
 Best Local Similarity 42.8%; Pred. No. 3.5e-204;
 Matches 494; Conservative 110; Mismatches 378; Indels 172; Gaps 17;

Qy 14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHFFDQIYYTSCTDILNW---FKCT 70
 ||| : || |||||: |||: | :||: || : || |
 Db 7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKESHLRPFSLPAESCH--RPWEDPHTCA 64

Qy 71 RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
 : | ||| || | : ||| :||| ||| || :|||:||| ||| |||
 Db 65 QPTVVYRTVYRQVVKMDSRRLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW 124

Qy 131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
 | :||| | ||| | | | : |:| :| | :| : | | ||
 Db 125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDDPKSGTCFCPSGLQPPNCLQPCPAGHYGPA 184

Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPQCEQRCPCQNGGVCHHVT 250
 | ||| ||:|| | |||| | | | | | : |||||
 Db 185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGDGFFCPRTYPCQNGGVPQGSQ 243

Qy 251 GECSCPSGWMGTVCGQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
 | |||| |||| :| |||| | ||:||:|||| || |||||:||| |:|||:
 Db 244 GSCSCPPGWMGVICSLPCPEGFHGPNTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

Qy 311 CPVGTYGVLCAETCQCVNNGGKCYHVSAGLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
 |||| :| |||| | | :|: :||| || |:|| ||||:| ||: | : | |
 Db 304 CPVGRFGQDCAETCDCAPGARCFFANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

Qy 371 HLENTSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
 |:| |||| |||:|:|:|:|:|:| :| ||: | | :| | : :| | |
 Db 364 DPEHSLSCHPMHGECSQPGWAGLHCNESCQDTHGPGCQEHCLCLHGGLCLADSGLCRC 423

Qy 431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDAVCSPVDGSCCTKAGWHGVDCSIRCPSTG 490
 |||: | | : || ||||| ||| :| |||:|:| ||| :||: || ||
 Db 424 APGYTGPHCANLCPDPTYGINCSSRCSCENAIACSPIDGTICKEGWQRGNCSVPCPLGT 483

Qy 491 WGFGCNLTQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCHADGC 550
 ||| || :|| : | | : | || || | | :|| | :| || ||| :|||
 Db 484 WGFNCNASCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKQGFEGGCASVCDCHSDGC 543

Qy 551 HPTTGHCRCPLGWSGVHCDVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTT 610
 | | ||| || | | | || || || | |||| :| :| | ||||| :
 Db 544 DPVHGQCRCQAGWMGTRCHLPCPEGFWGANCSNTCTCKNGGTCVSENGNCVCAPEGFRGPS 603

Qy 611 CQRICSPGFYGHRCSTCPQCQVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA 670

```

      ||| | || || || |
Db      604 CQRPCPPGRYGKRCVQ----- 619
Qy      671 GICTCTNN-GTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDG 729
      | | || :|| | :| | | | ||| : || | || | | | :| | |
Db      620 --CKCNNNHSSCHPSDGTCSCLAGWTGPDCSEACPPGHWGLKCSQLCQCHHGGTCHPQDG 677
Qy      730 ECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKC 789
      | ||||| | : || :| :| :|| | | | | |
Db      678 SCICTPGWTGPNCLEGCPFRMFGVNCSQLCQCDLGEMC-----HPE--- 718
Qy      790 PSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTAL 849
      | | | | | | | | | | | | | | | | | | :| : | : |
Db      719 -----TGACVCPPGHSGADCK-----MGSQESFTIMPTS- 747
Qy      850 PADSQIGAIAGIILVLVFLFLLALFIIYRHKQKGKESSMPAVTYTPAMRVVNADYTIS 909
      | :|| :| :| :| :||| | | |||| | | :| :| :| :
Db      748 PVTHNSLGAVIGIAVLGTLVVALIALFIGYRQWQKGKEHEHLAVAYSTG-RLDGS DYMP 806
Qy      910 GTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVNLKVNPNPK 969
      | : ||| :||| ||| :| :| | | :||| :
Db      807 DVSP-----SYSHYSNPSYHTLSQCSNP PPPN-----KVPGSQLFVSSQAPERPS 853
Qy      970 RGPVGDCGTGLPADWKHGGYLNELGAFGLDRSY-----MGKS 1006
      | : ||||| :| || |||| | | :| |
Db      854 RAHGRENHVTLPADWKHREPHERGASHLDRSYSCSYSHRNGPGPFCHKGPIS EELGAS 913
Qy      1007 LKDLGKNSEYNSSNCSLSSSENPYATIKDPPVLIPKSSECGYVEMKSPARRDSPYAEINN 1066
      : | ||||| :| | | : | |||| | | :
Db      914 VMSL-----SSENPYATIRDLPSLPGEPR ESGYVEMKGPPSVSPPRQSLH- 958
Qy      1067 STSANRNVYEVEP-----TVSVVQGVFSNNGRLSQDP-----YDLPKNSHIPCHY 1111
      :| :|| | : | | | | | ||||| ||
Db      959 --LRDRQQRQLQPQRDSGTYEQPSPLSHNEESLGSTPPLPPGLPPGHYDSPKNSHIPGHY 1016
Qy      1112 DLLPVRDSSSSPKQ 1125
      || || | | :
Db      1017 DLPPVRHPPSPPSR 1030

```

RESULT 8

Q8CGA7

ID Q8CGA7 PRELIMINARY; PRT; 1004 AA.

AC Q8CGA7;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Similar to RIKEN cDNA 3110045G13 gene.

GN 3110045G13RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RF SEQUENCE FROM N.A.

RC STRAIN=FVB/N;

RA Strausberg R.;

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC042490; AAH42490.1; -.
 DR MGD; MGI:1920432; 3110045G13Rik.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 6.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 14.
 DR SMART; SM00180; EGF_Lam; 12.
 DR PROSITE; PS00022; EGF_1; 12.
 DR PROSITE; PS01186; EGF_2; 11.
 SQ SEQUENCE 1004 AA; 107377 MW; 9508B0EC04561E94 CRC64;

Query Match 37.5%; Score 2529; DB 11; Length 1004;
 Best Local Similarity 41.3%; Pred. No. 3.6e-193;
 Matches 477; Conservative 105; Mismatches 368; Indels 206; Gaps 18;

Qy 14 LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSTCDILNW---FKCT 70
 ||| : || |||||: |||: | :||: || : || |
 Db 7 LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKEHLRPFSLPAESCH--RPWEDPHTCA 64

 Qy 71 RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW 130
 : | ||| || | : ||| :||| ||| || :|||:||| ||| |||
 Db 65 QPTVVYRTVYRQVVKMDSRPRLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCQCAPGW 124

 Qy 131 GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND 190
 | :||| | ||| | ||| : ||| :||| | :| : | | |||
 Db 125 RGGDCSSECAPGMWGPQCDKFCHCGNNSCDBPKSGACFCPSGLQPPNCLQPCPAGHYGPA 184

 Qy 191 CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPFGKHGPQCEQRCPCQNGGVCHHVT 250
 | ||| ||:|| | |||| | | | | | : |||||
 Db 185 CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGFFFCPRTYPCQNGGVPQGSQ 243

 Qy 251 GECSCPSGWMGTVCQGQPCPEGRFGKNCSSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDE 310
 | |||| |||| :| |||| | ||:||||:|||| || |||||:||| |:||:|
 Db 244 GSCSCPPGWMGVICSLPCPEGFHGPNTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE 303

 Qy 311 CPVGTYGVLCAETCQCVNNGKCYHVSAGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC 370
 |||| :| |||| | | :| : |||| | | :|| ||||:| ||: | : | |
 Db 304 CPVGRFGQDCAETCDCAFGARCFPANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC 363

 Qy 371 HLENTSCHPMMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTC 430
 |:: |||| | | :| | : :| | |
 Db 364 DPEHSLSCHPMH-----CLCLHGGGLCLADSGLCRC 393

 Qy 431 APGFKGIDCSTPCPLGTYGINCSSRCGCKNDVCSFVDGSCCTKAGWHGVDCSIRCPSTG 490
 |||: | | : || ||||| ||| :| |||:||:| || || :||: || ||
 Db 394 APGYTGPHCANLCPDPTYGINCSSRCSCENAIACSPIDGTICKEGWQRGNCVPCPLGT 453

 Qy 491 WGFGCNLTQCCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCHADGC 550
 ||| || :||| : | | : | ||| ||| | | :||| | :| || ||| |:|||
 Db 454 WGFNCNASCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFEGGCASVCDCHSDGC 513

 Qy 551 HPTTGHCRCPLGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPPDDGICECAPGFRGTT 610

	: :	
Db	514 DPVHGQCRCQAGWMGTRCHLPCPEGEFWGANCSNTCTCKNGGTCVSENGNCVCAPGFRGPS	573
Qy	611 CQRICSPGFYGHRSQTCPQCVHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA	670
Db	574 CQRPCPPGRYGKRCVQ-----	589
Qy	671 GICTCTNN-GTCNPIDRSCQCYPGWIGSDCSQCPPAHWGPNCIHTCNCHNGAFCSAYDG	729
	: : : :	
Db	590 --CKCNNNHSSCHPSDGTCSCLAGWTGPDCSEACPPGHWGLKCSQLCQCHHGGTCHPQDG	647
Qy	730 ECKCTPGWTGLYCTQRCPLGFYKDCALICQCQNGADCDHISGQCTCRTGFMGRHCEQKC	789
	: : : :	
Db	648 SCICTPGWTGPNCLEGCPPRMFGVNCSQLCQCDLGEMC-----HPE---	688
Qy	790 PSGTYGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTAL	849
	: : :	
Db	689 -----TGACVCPPGHSGADCK-----MGSQESFTIMPTS-	717
Qy	850 PADS YQIGAIAGIIILVLVLFLFLLALFIIYRHKKQGKESSMPAVTYTPAMRVVNADYTIS	909
	: : : : : : : : :	
Db	718 PVTHNSLGA VIGIAVLGLTLVVALIALFIGYRQWQKGEHEHLAVAYSTG-RLDGSDYVMP	776
Qy	910 GTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKS KNNQLFVNLKNVNP GK	969
	: : : : : : :	
Db	777 DVSP-----SYSHYYSNPSYHTLSQCSPNPPPPN-----KVPGSQLFVSSQAPERPS	823
Qy	970 RGPVGDC TGTLPADWKHGGYLNELGAFGLDRSY-----MGKS	1006
	: : :	
Db	824 RAHGRENHTTLPADWKHRREPHDRGASHLDRSYSCSYSHRNPGPFCHKGP ISEEGLGAS	883
Qy	1007 LKDLGKNSEYNSSNCSLSSSENPYATIKDPVLI PKSSSECYVEMKSPARRDSPYAEINN	1066
	: : : ::	
Db	884 VM SL-----SSEN PYATI RDLP SL PGEPRESGYVEMKGPPSVSPPRQSLH-	928
Qy	1067 STSANRNVYEVEPTVSVVQGVFSNNGRLSQDP-----YDLPKN SHIPC	1109
	: :: : :	
Db	929 --LRDRQQRQLQPQRD--SGTYEQPSPLSHNEESLGSTSPLPPGLPPGHYDSPKNSHIPG	984
Qy	1110 HYDLLPV RDSSSSPKQ	1125
	:	
Db	985 HYDLPPVRHPSPPSR	1000

RESULT 9

Q8VHF4

ID Q8VHF4 PRELIMINARY; PRT; 747 AA.

AC Q8VHF4;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Jedi-736 protein.

GN 3110045G13RIK.

OS *Mus musculus* (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI TaxID=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL; TISSUE=Testis;
 RA Krivtsov A.V., Zinovyeva M.V., Hendrikx J., Visser J.W.M.,
 RA Belyavsky A.V.;
 RT "Jedi is a novel DSL and EGF-like repeat motif-containing protein
 RT expressed on non-differentiated hematopoietic cells."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF461685; AAL66380.1; -.
 DR MGD; MGI:1920432; 3110045G13Rik.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR009030; Grow_fac_recep.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 6.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 4.
 DR PROSITE; PS00022; EGF_1; 13.
 DR PROSITE; PS01186; EGF_2; 12.
 KW EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 747 AA; 78972 MW; F825F8F384D4736A CRC64;

Query Match 34.0%; Score 2292.5; DB 11; Length 747;
 Best Local Similarity 48.5%; Pred. No. 1.9e-174;
 Matches 377; Conservative 75; Mismatches 272; Indels 53; Gaps 5;

Qy	14	LLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNW---FKCT	70
		: : :: : : :	
Db	7	LLLALGLRLTGTLNSNDPNVCTFWESFTTTTKESHLRPFSLPAESCH--RPWEDPHTCA	64
Qy	71	RHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTCQCEPGW	130
		: : : : :	
Db	65	QPTVVYRTVYRQVVKMSRRLQCCRGYYESRGACVPLCAQECVHGRCVAPNQCCAPGW	124
Qy	131	GGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGND	190
		: : : : : :	
Db	125	RGGDCSSECAPGMWGPQCDKFCHCGNNSCDDPKSGACFCPSGLQPPNCLQPCPAGHYGPA	184
Qy	191	CHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGGVCHHVT	250
		: :	
Db	185	CQFDCQCY-GASCDPQDGACFCPPGRAGPSCNVPCSQGTGDFFCPRTPYPCQNGGVPPQGSQ	243
Qy	251	GECSCPSGWMGTVCQPCPEGRFGKNCSEQCQCHNGGTCDAAATGQCHCSPGYTGERCQDE	310
		: : : : : :	
Db	244	GSCSCPPGWMGVICSLPCPEGFHGPNCQTQECRCHNGGLCDRFTGQCHCAPGYIGDRCQEE	303
Qy	311	CPVGTYGVLCAETCQCVNGGKCYHVS GACLCEAGFAGERCEARLCPEGLYGIKCDKRCPC	370
		: : : : : : :	
Db	304	CPVGRFGQDCAETCDCAPGARCFFANGACLCEHGFTGDRCTERLCPDGRYGLSCQEPCTC	363
Qy	371	HLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCD SVTGKCTC	430
		: : : : : : : : :	
Db	364	DPEHSLSCHPMHGECSQPGWAGLHCNESCQDTHGPGCQEHCLCLHGGGLCLADSGLCRC	423
Qy	431	APGFKGIDCSTPCPLGTYGINCSSRCGCKNDVCS PVDGSCTCKAGWHGVDCSIRCPST	490
		: : : : : : :	
Db	424	APGYTGPHCANLCPDPTYGINCSSRCSCENAIACSPIDGTICKEGWQRGNCSVPCPLGT	483

Qy 491 WGFGCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCHADGC 550
 ||| || :||| : | | : | ||| ||| | :||| | :| || ||| :|||
 Db 484 WGFNCNASCQCAHDGVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASVCDCHSDGC 543
 Qy 551 HPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTT 610
 | | ||| || | | | || ||| | |||| :| :|| | ||||| :
 Db 544 DPVHGQCRCQAGWMGTRCHLPCPEGFWGANCSTCTCKNGGTCVSENGNCVCAPGFRGPS 603
 Qy 611 CQRICSPGFYGHRCSTCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCA 670
 ||| | || || || |
 Db 604 CQRPCPPGRYGKRCVQ----- 619
 Qy 671 GICTCTNN~GTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDG 729
 | | || :||| | :| | || | ||| : ||| || | | ||| :| | ||
 Db 620 --CKCNNNHSSCHPSDGTCSCLAGWTGPDCEACPPGHWGLKCSQLCQCHHGGTCHPQDG 677
 Qy 730 ECKCTPGWTGLYCTQRCPLGFYKGKCALICQCNQADCDHISGQCTCRTGFMGRHCE 786
 | ||||| | : || :| :| :||| | | :| | | | | :
 Db 678 SCICTPGWTGPNCLEGCPPRMFGVNCSQLCQCDLGEMCHPQTGACVCPGHSADCK 734

RESULT 10

Q8ND91

ID Q8ND91 PRELIMINARY; PRT; 626 AA.
 AC Q8ND91;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN DKFZP434L121.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Poustka A., Wellenreuther R., Mewes H.W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDJB databases.
 DR EMBL; AL834326; CAD38994.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR005805; Rieske.
 DR Pfam; PF00008; EGF; 7.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 11.
 DR SMART; SM00180; EGF_Lam; 11.
 DR PROSITE; PS00022; EGF_1; 11.
 DR PROSITE; PS01186; EGF_2; 11.
 DR PROSITE; PS00200; RIESKE_2; 1.

KW Hypothetical protein; EGF-like domain; Laminin EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 626 AA; 64059 MW; C166FE1BD2A949F9 CRC64;

Query Match 31.7%; Score 2135.5; DB 4; Length 626;
Best Local Similarity 60.6%; Pred. No. 5.4e-162;
Matches 339; Conservative 61; Mismatches 144; Indels 15; Gaps 4;

QY 261 GTVCGQPCPEGRFGKNCSQECQCHNGGTCDAATGQCHCSPGYTGERCQDECPVGTYGVL 320
| || |||| | ||:||||:| ||:| | |||||: || |:||||:| ||:| |
Db 6 GAVCAQPCPPGTFGQNCSDCPCHGGQCDHVTGQCHCTAGYMGDRCQECPFGSFGFQC 65

QY 321 AETCQCVNGGKCYHVSAGLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCHP 380
:: | | |||:| :||| | | :| | | :| |||||:| | ||| :|| ||||
Db 66 SQRCDCCHNGGQCSPTTGACECEPGYKGRPCQERLCPEGLHGPCTLPCCDADNTISCHP 125

QY 381 MSGECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGFKGIDCS 440
::| | |:|||| :||:| |:|: | | |:||||| |:| | ||||| | |:
Db 126 VTGACTCQPGWSGHHCNESCPVGYGDCQLPCTCQNGADCHSITGGCTCAPGFMGEVCA 185

QY 441 TPCPLGTYGINCSSRCGCKNDAVCSPVDGSCCTCKAGWHGVDCSIRCPSTWGFGCNLTCQ 500
| |||| |||| | | | ||||| |||| | |:|:|: ||||| | |:
Db 186 VSCAAGTYGPNCSISCSNNGGTCSPVDGSCCTCKEGWQGLDCTLPCTSGTWGLNCNESCT 245

QY 501 CLNGGACNTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGHCRCL 560
| || ||: :||:|:| || | :| |||| | |:|:|:| | ||||| | |||| |
Db 246 CANGAACSPIDGSCSCTPGWLGDTCCLPCPDGTGFLNCSEHCDCSHADGCDPVTGHCCCL 305

QY 561 PGWSGVHCDVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTTTCQRICSPGFY 620
||:|: ||| | |||||:| |:| | |||:| | ||||| | |||| | ||||
Db 306 AGWTGIRCDSTCPPGRWGPNCVSCSCENGSGSPEDGSCECAPGFRGPLCQRICPPGFY 365

QY 621 GHRCSTQCPQCVHSSGPGCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGT 680
|| |:| || |||| | ||||:|:|:| ||||:| ||||:| | ||:|:| |:| | ||||
Db 366 GHGCAQPCPLCVHSSRPCHHISGICECLPGFSGALCNQVCAGGYFGQDCAQLCSCANNGT 425

QY 681 CNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGL 740
|:| | ||||:| |||| | ||| ||| | | |:| |||| | || | | ||||| |
Db 426 CSPIDGSCQCFPGWIGKDCSQACPPGFWGPACFHCACHNGASCSAEDGACHCTPGWTGL 485

QY 741 YCTQRCPLGFYKDCALICQCCQNGADCDHISGQCTCRTGFM---GRHCEQKCPSGTYGY 796
:| | | | : :| | | : | | | | | | | | | | |:
Db 486 FCTQRKPHLLASQPLRIPC-CGLLATVGIVQ---TSREGGMQAAPGLVVPDSCPTRTEEL 541

QY 797 GCRQICDCLNNSTCDHITG 815
| :| | | |
Db 542 -----CRGSSRPDWIQG 553

RESULT 11

O88281

ID O88281 PRELIMINARY; PRT; 1574 AA.

AC O88281;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE MEGF6.

GN MEGF6.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=98360089; PubMed=9693030;
 RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
 RT "Identification of high-molecular-weight proteins with multiple EGF-
 RT like motifs by motif-trap screening.";
 RL Genomics 51:27-34(1998).
 DR EMBL; AB011532; BAA32462.1; -.
 DR PIR; T13954; T13954.
 DR HSSP; P00736; 1APQ.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 20.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00179; EGF_CA; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 5.
 DR PROSITE; PS00022; EGF_1; 23.
 DR PROSITE; PS01186; EGF_2; 23.
 DR PROSITE; PS01187; EGF_CA; 5.
 KW EGF-like domain.
 SQ SEQUENCE 1574 AA; 165445 MW; 2B48533D8F77F6E7 CRC64;

Query Match 29.0%; Score 1958; DB 11; Length 1574;
 Best Local Similarity 41.3%; Pred. No. 2.7e-147;
 Matches 344; Conservative 77; Mismatches 306; Indels 106; Gaps 16;

QY 95 CP-GFYESGEMCVP--HCADKCVHGRC-IAPNTCQCEPGWGGTNCSSACDGDHWGPHCTS 150
 || ||| |: | |||:: ||| | :|| | | :|| |:
 Db 602 CPKGFY--GKHCRKKKCHCANR---GRCHRLYGACLCDPGLYGRFCHLACPPWAFGPGCSE 656
 QY 151 RCQCKNG--ALCNPI TGACHCAAGFRGWRCE DRCEQGTYGNDCHQRCQCQNGATCDHVTG 208
 | |: ||| |:| | |||:| ||: || | :| | || || | || |:|
 Db 657 DCLCEQSHTRSCNP KDGS CSCKAGFQGERCQAECESGFFGPGCRHRCTCQPGVACDPVSG 716
 QY 209 ECR--CPPGYTGAFCE DLCP PGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQG 266
 ||| |||| | | | | | | | | | | | | | | | | | | |
 Db 717 ECRTQCPPGYQGEDCGQEC PVGTFGVNCSGSCSCV-GAPCHRV TGECLCPPGKTGEDCGA 775
 QY 267 PCPEGRFGKNCSQEC-QCHNGGTCD AATGQCHCS PGYTGERCQDECPVGTYGVLCAETCQ 325
 |||||:| |: | |:| :|: || | | ||: | |||| | | || | |
 Db 776 DCPEGRWGLGCQEICPACEHGASCNPETGTCLCLPGFVGSRCQDTC SAGWYGTGCQIRCA 835
 QY 326 CVNGG-----KCYH-----VSGACLC 341
 | | | | | | | | | | | | | | | | | | | | | |
 Db 836 CANDGHCDPTTGRCSCAPGWTGLSCQRACDSGHWGPD CIHPCNCSAGHGNCDAVSG LCLC 895
 QY 342 EAGFAGERCEARLCPEGLYGIKCDKRCPC HLENTHSCHPMSGECACKPGWSGLYCNETCS 401

```

Db      896 EAGYEGPRCE-QSCRQGYGSPCEQKCR--EHGAACDHVSGACTCPAGWRGSFCEHACP 952
Qy      402 PGFYG-----EA-----CQQICSCQNG 418
      ||:| :| :| || |::| | :| :| :|| | | | :| |
Db      953 AGFFGLDCDSACNCSAGAPCDAVTGSCICPAGRWGPRCAQSCPPLTFGLNCSQICTCFNG 1012
Qy      419 ADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVAVCSPVDGSCCTCKAGWH 478
      | | | | | :| | | | :| | | | | | | | :| | :| | | |
Db      1013 ASCDSVTGQCHCAPGWMGPTCLQACPPGLYGKNCQHSCLCRNGGRCDPILGQCTCPEGWT 1072
Qy      479 GVDCSIRCPSTWGFNCNLTCQCLNGGACNTLDGTCTCAPGWRGEKCELPQDGTGYGLNC 538
      | :| | :| | | | | | | | | | | | | | :| | :| | :|
Db      1073 GLACENECLPGHYAAGCQLNCSCLHGGICDRLTGHCLCPAGWTGDKCQSSCVSGTFGVHC 1132
Qy      539 AERCDCSHADGCHPTTGHCRLPGWGSVHCDVCAEGRWGPNCSLPCYCKNGASCSPDDG 598
      | | | | | | | | | | | | | :| | :| | :| | | |
Db      1133 EEHCACRKGASCHHVTGACFCPPGWRGPHCEQACPRGWFGGEACAQRCLCPTNASCHHVTG 1192
Qy      599 ICECAPGFRGTTTCQRICSPGFYGHRCSTCPQCVHSSGPCHHITGLCDCLPGFTGALCNE 658
      | | | | | :| :| | | :| | | | :| | :| | | :| | :|
Db      1193 ECRCPPGFTGLSCEQACQPGTFGKDCEHLC-QCPGETWACDPASGVCTCAAGYHGTGCLQ 1251
Qy      659 VCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPNCIHTCNC 718
      | | | :| | | | | | | :| | | :| :| | | | :| :| | |
Db      1252 RCPSGRYGPGCEHICKCLNGGTCDPATGACYCPAGFLGADCSLACPQGRFGPSCAHVCAC 1311
Qy      719 HNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADCDHISGQCTCRT 778
      | | | | | :| | | | | :| | | | | :| | | :| | :|
Db      1312 RQGAACDPVSGACICSPGKTGVRCEHGCPODRFGKGCELKACACRNGGLCHATNGSCSCPL 1371
Qy      779 GFMGRHCEQKCPSTGYGCRQICDCLNNSTCDHITGTCYCSPGWKGARCDQA 831
      | :| | | | :| | | | | | | :| | | | | :| | :|
Db      1372 GWMGPHCEHACPAGRYGAACLLECFQNNGSCEPTTGACLCGPGFYGQACEHS 1424

```

RESULT 12

Q9TVQ2

```

ID   Q9TVQ2          PRELIMINARY;          PRT; 1664 AA.
AC   Q9TVQ2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Y64G10A.7 protein.
GN   Y64G10A.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mortimore B.J.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;

```

RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ainscough R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AL117206; CAB60454.1; -.
 DR EMBL; AL110498; CAB60454.1; JOINED.
 DR EMBL; AL110498; CAB57911.1; -.
 DR EMBL; AL117206; CAB57911.1; JOINED.
 DR HSSP; P00736; 1APQ.
 DR WormPep; Y64G10A.7; CE24549.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 22.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00179; EGF_CA; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; 22.
 DR PROSITE; PS01186; EGF_2; 24.
 DR PROSITE; PS01187; EGF_CA; 3.
 KW EGF-like domain.
 SQ SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;

Query Match 28.6%; Score 1931; DB 5; Length 1664;
 Best Local Similarity 39.2%; Pred. No. 4.1e-145;
 Matches 330; Conservative 85; Mismatches 311; Indels 116; Gaps 14;

Qy	95	CP-GFYESGEMCVPHCADKCVHGRCIAP-NTCQCEPGWGGTNCSSACDGDHWGPHCTSRC	152
		: : : :	
Db	664	CPDGFY--GSQCNLKRMDCPNGRCDPVFGYCTCPDGLYGQSCEKPCPHFTFGKNCRFP	721
Qy	153	QC--KNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGEC	210
		: : : : : : :	
Db	722	KCARENSEGDEITGKCRCKPGYGGHCKRMCSPGLEGAGCAMEKSCCPAGIRCDPVTGDC	781
Qy	211	--RCPPGYTGAFCELDLCPGKHPQCEQRCPC-----QNGGVCHHVTGECSCPSGWMGT	262
		: : : :	
Db	782	TKKCPAGYQGNLCDQPCPAGYFGYDCEQKCSADVAPSKSVCHHVTGTCTCLPGKTGP	841
Qy	263	VCGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAE	322
		: : : : : : : :	
Db	842	LCDQSCAPNTYGPNCATCSCVNGAKCDESDGSGCHCTPGFYGATCSEVCPTGRFGIDCMQ	901
Qy	323	TCQCVNCGKCYHVSAGLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMS	382
		: : : : : :	
Db	902	LCKCQNGAICDTSNGSCECAPGWSGKKCD-KACAPGTFGKDCSKKCDC-ADGMH-CDPSD	958
Qy	383	GECACKPGWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTGKCTCAPGFK-----	435
		: : : :	
Db	959	GECICPPGKKGHKCDETCDSGLFGAGCKGICSCQNGATCDSVTGSCECRPGWRGKKCDRP	1018

Qy	436	-----GIDCSTPCPLGTYG	449
		} :	
Db	1019	CPDGRFGEGCNAICDCTTTNDTSMYNPFVARCDHVTGECRCPAGWTGPDCQTSCLGRHG	1078
Qy	450	INCSSRCGCKNDAVCSVPDGSCTCKAGWHGVDCSIRCPSGTWGFGCNLTQCCLNGGACNT	509
		: : : :	
Db	1079	EGCRHSCQCSNGASCDRVTFGCDCPSGFMGKNCESECPEGLWGSNCMKHCLCMHGGECK	1138
Qy	510	LDGTCTCAPGWRGEKCELPQDGTYYGLNCAERCDCSHADGCHPTTGHCRCCLPGWSGVHCD	569
		: : : : :	
Db	1139	ENGDCCECIDGWTGPSCEFLCPFGQFGRNCAQRCNCKNGASCDRKTGRCECLPGWSGEHCE	1198
Qy	570	SVCAEGRWGPNCSLPCYCKNGASCSRDDGICECAPGFRGTTTCQRICSPGFYGHRCSTCP	629
		: : : :	
Db	1199	KSCVSGHYGAKCEETCECENGALCDPISGHCSCQPGWRGKKCNRPCLKGYFGRHCSQSC-	1257
Qy	630	QCVHSSGPGCHHITGLCDCLPGFTGALCNEVCP-----	661
		: : : :	
Db	1258	RCANSKS-CDHISGRCQCPKGYAGHSCTELCPDGTFGESCSQKCDGENSEMCDALSGKCF	1316
Qy	662	-----SGRFGKNCAGICTCTNNGTCNPFIDRSCQCYPGWIGSDCSQPCPPA	706
		: : : :	
Db	1317	CKPGHSGSDCKSGCVQGRFGPDCNQLCSCENGVCDSSTGSCVCPPGYIGTKCEIACQSD	1376
Qy	707	HWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDICALICQCQNGAD	766
		: : : : : : :	
Db	1377	RFGPTCEKICNCENGGTCDRLTGQCRCLPGFTGMTCNQVCPGFRFGAGCKEKCRCANG-H	1435
Qy	767	CDHISGQCTCRTGFMGRHCEQKCPSTYGYGCRQICDCLNNSTCDHITGTCTCYCSPGWKGA	826
		: : : :	
Db	1436	CNASSGECKCNLGTGPSCEQSCPSGKYGLNCTLDCECYGQARCDPVQGCCDCPPGRYGS	1495
Qy	827	RC 828	
Db	1496	RC 1497	

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
 RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
 RA Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
 RA Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
 RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
 RA Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AE003472; AAF47553.2; -.
 DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 7.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 12.
 DR SMART; SM00180; EGF_Lam; 11.
 DR PROSITE; PS00022; EGF_1; 11.
 DR PROSITE; PS01186; EGF_2; 13.
 DR PROSITE; PS00290; IG_MHC; 1.
 SQ SEQUENCE 881 AA; 96380 MW; 52196D164F52F5C1 CRC64;

Query Match 27.2%; Score 1832.5; DB 5; Length 881;
 Best Local Similarity 34.0%; Pred. No. 1.4e-137;
 Matches 342; Conservative 117; Mismatches 386; Indels 161; Gaps 17;

Qy 151 RCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATCDHVTGEC 210
 :| | | |:| | :| | | | :| | | | :| :| :| :|:|:| | | |:| |
 Db 2 QCDCLNNAVCEPFSGDCECAKGYTGARCADICPEGFFGANCKSEKCRCEGGKCHHVSSEC 61

Qy 211 RCPPGYTGAFCELDLCPGKHGPQCEQRCPCQNGGVCHHVTGECSCPSGWMGTVCQPCPE 270
 :| | |:| | :| | | | | | | | | | | | | | | | | | | | | | |
 Db 62 QCAPGFTGPLCDMRCPDGKHAQCQQDCPCQNDGKCQPETGACMCNPGWTGDVCAKNCPCV 121

Qy 271 GRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTGERCQDECPVGTYGVLCAETCQCVNGG 330
 | :| | :| |:| | | | | | | | | | | | | | | | | :| | | | |
 Db 122 GSYGPGCQESCECYKGAPCHHITGQCECPPGYRGERCFCDECQLNTYGFNCMTCDCCANDA 181

Qy 331 KCYHVSGACLCEAGFAGERCEARLCPEGLYGIKCDKRCPCCHLENTHSCHPMMSGECACKPG 390
 | :| | |:| | :| | :| | :| | | | | | | | | | | | | | | | |
 Db 182 MCDRANGTCICNPGWTGAKCAERICEANKYGLDCNRTCECDMEHTDLCHPETGNCQCSIG 241

Qy 391 WSGLYCNETCSPGFYGEACQQICSCQNGADCDSDVTGKCTCAPGFKGIDCSTPCPLGTYGI 450
 | | | | | :| | | | :| |:|:| | | | | | | | | | | | | | | | |
 Db 242 WSSAQCTRPCTFLRYGPNCELTNCNKGAKCSPVNGTCLCAPGWRGPTCEESCEPPTFGQ 301

Qy 451 NCSSRCGCKNDAVCSVPDGSCTCKAGWHGVDCSIRCPSTGWGFGCNLTQCLNGGACNTL 510
 |:| | | |:| | | | | | | | | | | | | | | | | | | | | | | |
 Db 302 DCALRCDCQNGAKCEPETGQCLCTAGWKNIKCDRCDLNHFQDCAKVCDCHNNAACNPQ 361

Qy 511 DGTCTCAPGWRGEKCELPQDGTYGLNCAERCDC--SHADGCHPTTGHCRCLPGWSGVHC 568
 |:|:| | | | | |:| | | | :| :|:|:| | | | | | | | | | | | |
 Db 362 NGSCTCAAGWTGERCERKCDTGKFGHDCAQKCQCDFNNSLACDATNGRCVCKQDW-GV-- 418

Qy 569 DSVCAEGRWGPNCSLPCYCKNGASCSPDDGICECAPGFRGTTTCQRICSPGFYGHRCSTC 628
 | | | | :| | | | | | | | | | | | | | | | | | | | | | | |
 Db 419 -----CRCLNNSCDPDSGNCICSAGWTGADCAEPCPPGFYGMCKERC 462

Qy 629 PQC VHSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSC 688
 | :| :| :| | | | | | | | | | | | | | | | | | | | | | | | |
 Db 463 PEILHGKSCDHITGEILCRTGYIGLTCEHPCPAGLYGPGCKLKCNCEHGGECNHVTGQC 522

Qy 689 QCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGECKCTPGWTGLYCTQRCPL 748
 | | | | | |:|:| | | | :| | | | | | | | | | | | | | | | |
 Db 523 QCLPGWTGSNCSNESCPTDTYQGCAQRCRCVHHKVKCRKADGMCICETGWSGTRCDEVCP 582

Qy	749	GFYGKDCALICQCQNGADCDHISGQCTCRITGFMGRHCEQKCP	SGTGYGCRQICDC	LNNS	808
		: : : : : :			
Db	583	GFYGEHCMNTCACPSANFQCHAAHGCVCRSGYTGDNCELIAS	-----QRIADQ	SENS	635
Qy	809	TCDHITGTCYCSFGWKGARCDQAGVIIVGNLNSLSRTSTALPADSYQIGAIAGIIILVLV			868
				:::	
Db	636	-----SRASVALT	-----LVLMTLF		650
Qy	869	VLFLLLALFIIYRHKQKGKESSMPAVTYTPAMRVVNADYTISGTLPHSNGGNANSHYFTNP			928
		: : : : :			
Db	651	ACIIFAVFIYYRRRVSNLKTEIAHVHYT	-----HDTNPFSWPPN	---HNFDNP	695
Qy	929	SYHTLTQCATS	PH----VNNRDRMTVTKSKNNQLFVNLKVNPNPGKRGFVGDC	-----TG	978
		: : : : : :			
Db	696	VYGMQAE	TRLPLNNMRSKMNNFDQRSTMSTDYGD	-----DCNASGRVG	738
Qy	979	TLPADWKHGGYLNELGAFLDRSYMGKSLKDLGKNSEYNSSNCSLSSSEN	PYATI----		1033
		: : : : : : :			
Db	739	SYSINYNHDLTLTKNLNADRTNP	IVYNESLKE-----EHVYDEIKHKEG		781
Qy	1034	-KDP-----PVLIPKSSECGYVEMKSPARRDSPYAEINNSTSANRNVYEVEPTVSVVQGV			1087
		: : : : : : :			
Db	782	YKDPVKIYSKILFPE-DEYDHL	DYSRPSTSQKPHYHRMNDAMLNINQDEEKPSNVKNMTV		840
Qy	1088	FSNNGRLSQDPYDLPKNSHIPCH	-----YDLLPVRDSSSSPK		1124
		:			
Db	841	LLNK-----PLPPTPEPEQHECFDNTNTNLDNVSTAS	PSSSPK		878

DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 4.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 7.
 DR PROSITE; PS00022; EGF_1; 15.
 DR PROSITE; PS01186; EGF_2; 11.
 KW EGF-like domain.
 SQ SEQUENCE 1070 AA; 114180 MW; 75254D0DD5643AE5 CRC64;

Query Match 27.0%; Score 1823.5; DB 5; Length 1070;
 Best Local Similarity 32.4%; Pred. No. 9.1e-137;
 Matches 362; Conservative 155; Mismatches 416; Indels 183; Gaps 34;

Qy 21 GTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYT-----SCTDILNWFKCTRHR 73
 || | : :|| : : | : : | : | :|
 Db 35 GTTEP---QGDHVCT----VKTIVDDY--ELKKVIHTVVYNDTEQCLNPLTGFQC---- 80
 Qy 74 VSYRTAYRHGEKTMYYRK-----SQCCPGFYESGE-MCVPHCADKCVHGRCIAPNTC 124
 | : ||| | :| : ||| | :| : | : ||| | |
 Db 81 ----TVEKRGQKASYQRQLVKKEKYVKQCCDGYQTKDHFCLPDCNPPCKKGKCIIEPGKC 136
 Qy 125 QCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQ 184
 : ||| : || : :| : | : ||| : | | : ||| : ||| |
 Db 137 ECDPGYGGKYCASSCSVGTWGLGCSKSCDCENGANCDPELGTCICTSGFQGERCEKPCPD 196
 Qy 185 GTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGG 244
 : | : | : | |||| | : | : | | : | | | : | : | ||||
 Db 197 NKWGPNCVKSCPCQNGGKCNK-EGKCVCSDBGWGEFCLNKCEEKGFGAECKFECNCQNGA 255
 Qy 245 VCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATQCHCSPGYTG 304
 | : | : | | || : | | | || | : | | | : | | | : | : |
 Db 256 TCDNTNGKCICKSGYHGALCENECVSGFFGSGCTQKCDCLNNQNCSSSGECKCI-GWTG 314
 Qy 305 ERCQDECPVGTYGVLCAETCQCV-----NGGKCYHVSGACLCEAGFAGERCEARLCPEG 358
 : | | | : : | : | | : | : | | : | : | : | : |
 Db 315 KHCIDIGCSRGRFGLQCKQNCCTCPGLEFSDSNASCDAKTGQCQCESGYKGPCKDERKCDAE 374
 Qy 359 LYGIKCDKRCPCCHLENTHSCHPMMSGECACKPGWSGLYCNETCSPGFYGEAC--QQICSCQ 416
 || | | | | || | | : | | ||| : | | || | | : |
 Db 375 QYGADCSKTCTCVRENTLMCAPNTGFRCRCKPGFYGDNCELACSKDSYGPNCEKQAMCDWN 434
 Qy 417 NGADCDSVTGKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAV-CSPVDGSCTCKA 475
 : : || : || | || | : || |||| || || : | | | | | |
 Db 435 HASECNPETGSCVCKPGRGTGKNCSEPCPLDFYGPNAHQQCNCQRGVCGDADGKCQCDR 494
 Qy 476 GWHGVDCSIRCPSGTWGFGCNLTQCQLNGGACNTLDGTCTCAPGWRGEKCELPCQDGTYG 535
 || | | || : || | : | | : : | || | : | : : | : ||
 Db 495 GWTGHRCEHHCPADTFGANCKRCKCPKGIGCDPITGECTCPAGLQGANCDIGCEPSYG 554
 Qy 536 LNCAERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSP 595
 | | | : | | | | || : | | : : : : | : | | | : ||||
 Db 555 PGCKLHCKCVNGK-CDKETGECTCQPGFFGSDCSTTCSKGKYGESCELSGPCSD-ASCSK 612
 Qy 596 DDGICECAPGFRGTTQQRICSPGFYGHRCSTQCPQCVHSSGPCHHI---TGLC-DCLPGF 651
 | | | | : | : | : | | : | : | : | | : | | |
 Db 613 QTGKCLCPLGTGKGVSCDQKCDPNTFGFLCQETV-----TPSPCASTDPKNGVCLSCPPGS 667

Qy 652 TGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSCQCYPGWIGSDCSQPCPPAHWGPN 711
 :| | |::| :| | :||: | :| | |::| |::| | :|
 Db 668 SGIHCEHNCPAGSYGDGCQQVCSCADGHGCDPTTGECICEPGYHGKTCSEKCPDGKYGYG 727
 Qy 712 CIHTC-NCHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYGKDCALICQCQNG-ADCDH 769
 | | |::| :| | | | |::| |::| | :||: | :|
 Db 728 CALDCPKCASGSTCDHINGLCICPAGLEGALCTRPCSAGFWGNGCRQVCRCTSEYKQCNA 787
 Qy 770 ISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCLNNST--CDHITGTCYCSFGWKGAR 827
 :||::| | | |::| | | | :| | :::| |::| |::|
 Db 788 QTGECSCPAGFQGDRC DKPCEDGYYGPD CIKKCKCQGTATSSCNRVSGACHCHPGFTGEF 847
 Qy 828 CDQAGVIIVGNLNSLSRTST-----ALP-----ADSYQIGAIAG----- 861
 | :|| | | | | :| | |
 Db 848 C-----HALCPESTFGLKCSKECPKDGCGDGYECDAAIGCCHVDQMSCGKAKQE 896
 Qy 862 -----IIILVLVVLVLF----LLALFIIYRHK-QKGKESSMPAVTYTPAMRV 901
 :||:| | |::| |::| | | :| |::|
 Db 897 FEALNGAGRSTGLTWFFVLLIVALCGGLGLIALF--YRNKYQKEKDPDMPTVSF----- 948
 Qy 902 VNADYTISGTLPHSNGGNANSHYFTNPSYHTLTQCATSPHVNNRDRMTVTKSKNNQLFVN 961
 | | | | : : | : | : :||
 Db 949 -----HKAPNNDEGREFQNPPLY---SRQSVFP---DSDAFSSENNGNHQ---- 986
 Qy 962 LKNVNPGRGPVGDCTGTLPADWK-----HGGYLNELGAFGLDRSYMKG-----SLK 1008
 | | | : : | | | | : | |
 Db 987 -----GGPPNGLLTLEEEELNKKIHG-----RSAAGRGNNDYASLD 1023
 Qy 1009 DLGKNSEYNSSNCSLSSSENPYATIKDPPVLIPKSS 1044
 : : :||: | | |||| | | : |:
 Db 1024 EVAGEGSSSSASASASRRENPYADISSDPVTQNSA 1059

RESULT 15

Q9XWD6

ID Q9XWD6 PRELIMINARY; PRT; 1111 AA.
 AC Q9XWD6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Y47H9C.4 protein (CED-1).
 GN Y47H9C.4 OR CED-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 RT *elegans*.";
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21097720; PubMed=11163239;
 RA Zhou Z., Hartwig E., Horvitz H.R.;
 RT "CED-1 is a Transmembrane Receptor that Mediates Cell Corpse
 RT Engulfment in *C. elegans*.";
 RL Cell 104:43-56(2001).
 DR EMBL; AL032657; CAA21739.1; -.
 DR EMBL; AF332568; AAG60061.1; -.
 DR PIR; T26972; T26972.
 DR HSSP; P05106; 1JV2.
 DR WormPep; Y47H9C.4a; CE20264.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006212; Furin_repeat.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 4.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 6.
 DR SMART; SM00261; FU; 2.
 DR PROSITE; PS00022; EGF_1; 15.
 DR PROSITE; PS01186; EGF_2; 11.
 KW EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 1111 AA; 118803 MW; A39F374C008F9874 CRC64;

Query Match 26.8%; Score 1805.5; DB 5; Length 1111;
 Best Local Similarity 31.8%; Pred. No. 2.6e-135;
 Matches 373; Conservative 162; Mismatches 423; Indels 215; Gaps 40;

QY 21 GTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYT-----SCTDILNWFKCTRHR 73
 || | : :|| : : | : : | : | :|
 Db 35 GTTEP---QGDHVCT-----VKTIVDDY--ELKKVIHTVVYNDTEQCLNPLTGFQC--- 80
 QY 74 VSYRTAYRHGEKTMYYRRK-----SQCCPGFYESGE-MCVPHCADKCVHGRCIAPNTC 124
 | : || | :|| || | : : : | : | | :|| | |
 Db 81 ----TVEKRGQKASYQRQLVKKEKYVKQCCDGYQTGDHFLPDCNPPCKKGKCIEPGKC 136
 QY 125 QCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGALCNPITGACHCAAGFRGWRCEDRCEQ 184
 :||:|:| | :||:| || | : | | :||| | :| | | :||:| ||| |
 Db 137 ECDPGYGGKYCASSCSVGTWGLGCSKSCDCENGANCDPELGTCICTSGFQGERCEKPCPD 196
 QY 185 GTYGNDCHQRCQCQNGATCDHVTGECRCPPGYTGAFCEDLCPGKGHPQCEQRCPCQNGG 244
 :| :| : | |||| | : | :| | | : | || | :| : | ||||
 Db 197 NKWGPNCVKSCPCQNGGKCNK-EGKCVCSDBGWGGEFCLNKCEEKGKFGAECKFECNCQNGA 255
 QY 245 VCHHVTGECSCPSGWMGTVCQGQPCPEGRFGKNCSQECQCHNGGTCDAAATGQCHCSPGYTG 304
 | : | :| | ||: | :| | | || | :|| | | | :||:| | | :||
 Db 256 TCDNTNGKCICKSGYHGALCENECSVGGFFGSGCTQKCDCLNNQNCSSSGECKCI-GWTG 314

Qy	305	ERCQDECPVGTYGVLCAETCQCV-----NGGKCYHVS	GACLCEAGFAGERCEARLCPEG	358
Db	315	KHCDIGCSRGRFGLQCKQNCTCPGLEFSDSNASCDAKTGCQC	QCESGYKGPCKDERKCDAE	374
Qy	359	LYGIKCDKRCPCHLENTHSCHPMSGECACKPGWSGLYCNET	CSPGFYGEAC--QQICSCQ	416
Db	375	QYGADCSKTCTCVRENTLMCAPNTGFCRCCKPGFYGDNCEL	ACSKDSYGPNCCEKQAMCDWN	434
Qy	417	NGADCDSVTGKCTCAPGFGKIDCSTPCPLGTYGINCSSRC	GCKNDVAV-CSPVDGSGCTCKA	475
Db	435	HASECNPETGSCVCKPGRTGKNCSEPCPLDFYGPNC	AHQCCQNQRGVGCDGADGKCQCDR	494
Qy	476	GWHGVDCSIRCPSGTWGFGCNLTCQCLNGGACNTLDGTCT	CAPGWRGEKCELPQDGTYG	535
Db	495	GWTGHRCEHHCPADTFGANCKRKCKPKGIGCDPITGECT	CPAGLQGANCDIGCPEGSYG	554
Qy	536	LNCAERCDCSHADGCHPTTGHCRLPGWSGVHCDSVCAE	GRWGPNCSLPCYCKNGASCSP	595
Db	555	PGCKLHCKCVNGK-CDKETGECTCQPGFFGSDCSTT	SKGKYGESCELSGPCSD-ASCSK	612
Qy	596	DDGICECAPGFRGTTTCQRICSPGFYGHRC	SQTCPQCVHSSGPCHHI---TGLC-DCLPGF	651
Db	613	QTGKCLCPLGTGKVSCDQKCDPNTFGFLCQETV-----	TPSPCASTDPKNGVCLSCPPGS	667
Qy	652	TGALCNEVCPSGRFGKNCAGICTCTNNGTCNPIDRSC	QCYPGWIGSDCSQPCPPAHWGPN	711
Db	668	SGIHCEHNCPAGSYGDGCQQVCSCADGHGCDPTTGEC	ICEPGYHGKTCSEKCPDGKYGYG	727
Qy	712	CIHTC-NCHNGAFCSAYDGECKCTPGWTGLYCTQRC	PLGFYGKDICALICQCQNG-ADCDH	769
Db	728	CALDCPKCAGSTCDHINGLCICPAGLEGALCTRPC	SAGFWNGCRQVCRCTSEYKQCNA	787
Qy	770	ISGQCTCRTGFMGRHCEQKCPSGTYGYGCRQICDCL	NNST--CDHITGTCYCSPGWKGAR	827
Db	788	QTGECSCPAGFQGDRCDKPCEDGYYGPD	CIKKCKCQGTATSSCNRVSGACHCHPGFTGEF	847
Qy	828	CDQAGVIIIVGNLNSLSRTST-----ALP-----	ADSYQIGAIAG-----	861
Db	848	C-----HALCPESTFGLKCSKECPKDGCGDGYE	CDAAGCCHVDQMSCGKAKQE	896
Qy	862	-----IIILVLVLF-----LLALFIIYRHK-Q	KGKESSMPAVTYTPAMRV	901
Db	897	FEALNGAGRSTGLTWFFVLLIVALCGGLGLIALF--	YRNKYQKEKDPDMPTVSF-----	948
Qy	902	VNADYTIISGTLPHSNGGNANSHYFTNPSYHTLTQ	CATSEPHVNNRDRMTVTKSKNNQLFVN	961
Db	949	-----HKAPNNDEGREFQNPPLY--SRQSVFP--	DSDAFSSENNNGNHQ----	986
Qy	962	LKNVNPGRGPVGDCTGTLPADWKHGGYLNELGAFGL	DRSYM-----GKSLKDLGKN--	1013
Db	987	-----GGPPN--GLLTLEEEELENKKIHGR	SAAGRGNNDY	1019
Qy	1014	-----SEYNSSNCSLSSE--NPYATIKDPPVL	IPKSSSECGYVEMK----SPA----	1055
Db	1020	ASLDEVAGEGSSSSASASARRGLNSSEOSRRP--	LLEEHDEEEFDEPHENSISPAAV	1077

```
Qy      1056 ---RRDSPYAEINN-----STSANR----NVY 1075
          ::|||:|::      |||:      |:|
Db      1078 TSNHNENPYADISSDPVTONSANKKRAQDNLY 1110
```

Search completed: March 26, 2004, 16:11:10
Job time : 61.4809 secs

OM protein - protein search, using sw model

Run on: March 26, 2004, 15:58:50 ; Search time 19.1541 Seconds
(without alignments)
3099.072 Million cell updates/sec

Title: US-10-092-390-2
Perfect score: 6744
Sequence: 1 MVISLNSCLSFICLLCHWI.....SSPKQEDSGGSSSNSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	1037	15.4	2524	1	NOTC_XENLA	P21783 xenopus lae
2	1034.5	15.3	2556	1	NTC1_HUMAN	P46531 homo sapien
3	1028	15.2	2531	1	NTC1_MOUSE	Q01705 mus musculu
4	1024	15.2	2531	1	NTC1_RAT	Q07008 rattus norv
5	1014.5	15.0	2471	1	NTC2_HUMAN	Q04721 homo sapien
6	998	14.8	2471	1	NTC2_RAT	Q9qw30 rattus norv
7	993	14.7	2470	1	NTC2_MOUSE	O35516 mus musculu
8	987	14.6	2437	1	NTC1_BRARE	P46530 brachydanio
9	978.5	14.5	2703	1	NOTC_DROME	P07207 drosophila
10	977.5	14.5	2318	1	NTC3_MOUSE	Q61982 mus musculu
11	974	14.4	2321	1	NTC3_HUMAN	Q9um47 homo sapien
12	969.5	14.4	2319	1	NTC3_RAT	Q9r172 rattus norv
13	959.5	14.2	2003	1	NTC4_HUMAN	Q99466 homo sapien
14	954.5	14.2	1064	1	FBP1_STRPU	P10079 strongyloce
15	951.5	14.1	1964	1	NTC4_MOUSE	P31695 mus musculu
16	916.5	13.6	4289	1	TENX_HUMAN	P22105 homo sapien
17	870.5	12.9	830	1	SREC_HUMAN	Q14162 homo sapien

18	832.5	12.3	1213	1	JAG3_BRARE	Q90y54	brachydanio
19	813	12.1	833	1	SRC2_MOUSE	P59222	mus musculu
20	808	12.0	870	1	SRC2_HUMAN	Q96gp6	homo sapien
21	789	11.7	1238	1	JAG2_HUMAN	Q9y219	homo sapien
22	775.5	11.5	2139	1	CRB_DROME	P10040	drosophila
23	775	11.5	1242	1	JAG1_BRARE	Q90y57	brachydanio
24	769	11.4	2201	1	TENA_HUMAN	P24821	homo sapien
25	768	11.4	1247	1	JAG2_MOUSE	Q9qye5	mus musculu
26	767.5	11.4	1746	1	TENA_PIG	Q29116	sus scrofa
27	757	11.2	1218	1	JAG1_HUMAN	P78504	homo sapien
28	745	11.0	1202	1	JAG2_RAT	P97607	rattus norv
29	744	11.0	1219	1	JAG1_RAT	Q63722	rattus norv
30	739	11.0	1218	1	JAG1_MOUSE	Q9qxx0	mus musculu
31	736.5	10.9	3695	1	LMA5_HUMAN	O15230	homo sapien
32	720.5	10.7	1808	1	TENA_CHICK	P10039	gallus gall
33	717.5	10.6	1801	1	LMB2_RAT	P15800	rattus norv
34	716	10.6	3106	1	LMA2_MOUSE	Q60675	mus musculu
35	706.5	10.5	1790	1	LMB1_DROME	P11046	drosophila
36	704.5	10.4	3084	1	LMA1_MOUSE	P19137	mus musculu
37	700.5	10.4	1799	1	LMB2_MOUSE	Q61292	mus musculu
38	692	10.3	3075	1	LMA1_HUMAN	P25391	homo sapien
39	690	10.2	1408	1	SERR_DROME	P18168	drosophila
40	687.5	10.2	3110	1	LMA2_HUMAN	P24043	homo sapien
41	685.5	10.2	1798	1	LMB2_HUMAN	P55268	homo sapien
42	683.5	10.1	3718	1	LMA5_MOUSE	Q61001	mus musculu
43	676.5	10.0	3672	1	LML2_CAEEL	Q21313	caenorhabdi
44	667	9.9	1786	1	LMB1_HUMAN	P07942	homo sapien
45	664	9.8	1786	1	LMB1_MOUSE	P02469	mus musculu

ALIGNMENTS

RESULT 1

NOTC_XENLA

ID NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein).
GN XOTCH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch."
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M33874; AAB02039.1; -.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 29.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
 FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1729 1750 POTENTIAL.
 FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 57 EGF-LIKE 1.
 FT DOMAIN 58 99 EGF-LIKE 2.
 FT DOMAIN 102 140 EGF-LIKE 3.
 FT DOMAIN 141 177 EGF-LIKE 4.
 FT DOMAIN 179 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 217 254 EGF-LIKE 6.
 FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	294	332	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	334	370	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	371	409	EGF-LIKE 10.
FT	DOMAIN	411	449	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	451	487	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	489	525	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	527	563	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	565	600	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	602	638	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	640	675	EGF-LIKE 17.
FT	DOMAIN	677	713	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	715	750	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	752	788	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	790	826	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	828	866	EGF-LIKE 22.
FT	DOMAIN	868	904	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	906	942	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	944	980	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	982	1018	EGF-LIKE 26.
FT	DOMAIN	1020	1056	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1058	1094	EGF-LIKE 28.
FT	DOMAIN	1096	1142	EGF-LIKE 29.
FT	DOMAIN	1144	1180	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1182	1218	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1220	1264	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1266	1304	EGF-LIKE 33.
FT	DOMAIN	1306	1346	EGF-LIKE 34.
FT	DOMAIN	1347	1383	EGF-LIKE 35.
FT	DOMAIN	1386	1424	EGF-LIKE 36.
FT	REPEAT	1441	1478	LIN/NOTCH 1.
FT	REPEAT	1479	1520	LIN/NOTCH 2.
FT	REPEAT	1521	1560	LIN/NOTCH 3.
FT	REPEAT	1876	1919	ANK 1.
FT	REPEAT	1924	1953	ANK 2.
FT	REPEAT	1957	1987	ANK 3.
FT	REPEAT	1991	2020	ANK 4.
FT	REPEAT	2024	2053	ANK 5.
FT	REPEAT	2057	2086	ANK 6.
FT	DISULFID	22	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.
FT	DISULFID	167	176	BY SIMILARITY.
FT	DISULFID	183	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.

FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	504	BY SIMILARITY.
FT	DISULFID	498	513	BY SIMILARITY.
FT	DISULFID	515	524	BY SIMILARITY.
FT	DISULFID	531	542	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	553	562	BY SIMILARITY.
FT	DISULFID	569	579	BY SIMILARITY.
FT	DISULFID	574	588	BY SIMILARITY.
FT	DISULFID	590	599	BY SIMILARITY.
FT	DISULFID	606	617	BY SIMILARITY.
FT	DISULFID	611	626	BY SIMILARITY.
FT	DISULFID	628	637	BY SIMILARITY.
FT	DISULFID	644	654	BY SIMILARITY.
FT	DISULFID	649	663	BY SIMILARITY.
FT	DISULFID	665	674	BY SIMILARITY.
FT	DISULFID	681	692	BY SIMILARITY.
FT	DISULFID	686	701	BY SIMILARITY.
FT	DISULFID	703	712	BY SIMILARITY.
FT	DISULFID	719	729	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	756	767	BY SIMILARITY.
FT	DISULFID	761	776	BY SIMILARITY.
FT	DISULFID	778	787	BY SIMILARITY.
FT	DISULFID	794	805	BY SIMILARITY.
FT	DISULFID	799	814	BY SIMILARITY.
FT	DISULFID	816	825	BY SIMILARITY.
FT	DISULFID	832	843	BY SIMILARITY.
FT	DISULFID	837	854	BY SIMILARITY.
FT	DISULFID	856	865	BY SIMILARITY.
FT	DISULFID	872	883	BY SIMILARITY.
FT	DISULFID	877	892	BY SIMILARITY.
FT	DISULFID	894	903	BY SIMILARITY.
FT	DISULFID	910	921	BY SIMILARITY.
FT	DISULFID	915	930	BY SIMILARITY.
FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	986	997	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	1024	1035	BY SIMILARITY.

FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	1046	1055	BY SIMILARITY.
FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	1100	1121	BY SIMILARITY.
FT	DISULFID	1115	1130	BY SIMILARITY.

Query Match 15.4%; Score 1037; DB 1; Length 2524;
 Best Local Similarity 25.4%; Pred. No. 2.2e-54;
 Matches 326; Conservative 84; Mismatches 304; Indels 568; Gaps 78;

Qy	83	GEKTMYYR-----KSQC-----CP-GFYESGEMCVPHCADKCVHGR	117
		: : : : : :	
Db	53	GERCQFPNPCTIKNQCMNFGTCEPVLQGNADFICHCPVGF--TDKVCLTPVDNACVNNP	110
Qy	118	C-----IAPNTCQCEPGWGGTNCSSACDGDHGWPHCTSRCQCKNGALCNP--IT	164
		: :	
Db	111	CRNGGTCELLNSVTEYKCRCPPGWTGDSCQQA-----DPCASN-PCANGGKCLPFEIQ	162
Qy	165	GACHCAAGFRGWRCE---DRCEQ-----GTYGNDCHQR-----CQ	196
		: : :	
Db	163	YICKCPPPGFHGATCKQDINECSQNPCKNGGQCINEFGSYRCTCQNRFTGRNCDEPYVPCN	222
Qy	197	---CQNGATC---DHVTGECRCPPGYTGAFCE-----LC	225
		: : :	
Db	223	PSPCLNGGTCRQTDTSYDCTCLPGFSGQNCENIDDCPSNNCRNGGTCVDGVNTYNCQC	282
Qy	226	PPGKHGPQCEQ---RC-----PCQNGGVCHHVTG--ECSCPSGWMGTVCQ-----	266
		: : :	
Db	283	PPDWTGQYCTEDVDECQLMPNACQNGGTCHNTYGGYNCVCVNGWTGEDCSENIDDCANAA	342
Qy	267	-----PCPEGRFGKNC--SQEC---QCHNGGTCD--ATGQ---CHCSPG	301
		: :	
Db	343	CHSGATCHDRVASFYCECPHGRTGLLCHLDNACISNPCNEGSNCDTNPVNGKAICTCPPG	402
Qy	302	YTGERCQ---DECPVGTYGVLC AETCQCVNGGKCYHVSGA--CLCEAGFAGERCEARLCP	356
		: : : : :	
Db	403	YTGPACNNDVDECSLGN-----PCEHGGRCTNTLGSFQCNCPPQGYAGPRCEIDV--	452
Qy	357	EGLYGIKCDKRC---PCHLENTHSCHPMSGE--CAKPGWSGLYC-----	396
		: :	
Db	453	-----NECLSNPC--QNDSTCLDQIGEFQCICMPGYEGLYCETNIDECASNPCLN	501
Qy	397	-----NE---TCSPGFYGEACQ---QICS---CQNGADC-----DSVTGK-	427
		: : :	
Db	502	GKCIDKINEFRDCPTGFSGNLCQHD FDECTSTPCKNGAKCLDGPNSYTCQCTEGFTGRH	561
Qy	428	-----CTCAPGFKGIDC-----STP-----	442
		:	
Db	562	CEQDINECIPDPCHYGTCKDGIATFTCLCRPGYTGRLCDNDINECLSKPCLNGGQCTDRE	621
Qy	443	-----CPLGTYGINCSSR---CG---CKNDAVCSPVDG-SCTCKAGWHGVDCSIR----	485
		: : : :	
Db	622	NGYICTCPKGTGTVNCETKIDDCASNLC DNGKCIDKIDGYECTCEPGYTGLKCNININEC	681
Qy	486	-----CPSGTWGFGC---	495

Db	682	DSNPCRNGGTCKDQINGFTCVCPDGYHDMCLSEVNECNSNPCIHGACHDGVNGYKCDCE	741
Qy	496	-----NLTCQ---CLNGGACNTLDGT--CTCAPGWRGEKCEL-----PC-Q	530
Db	742	AGWSGSNCDINNNECESNPCMNGGTCKDMTGAYICTCKAGFSGPNCQTNINECSSNPCLN	801
Qy	531	DGT-----YGLNC-----AERCD-----CSHADGCHPT----TGHCRCCLPGWS	564
Db	802	HGTCIDDVAGYKCNCLPYTGAICEAVLAPCAGSPCKNGGRCKESEDFFETFSCECPPGWQ	861
Qy	565	GVHCD-----SVCAEGRWGPNC SL-----PCYCKNGASC	593
Db	862	GQTCEIDMNECVNRPCRNGATCQNTNGSYKCNCKPGYTGRNCEMDIDDCQPNPCHNGGSC	921
Qy	594	SPDDGI----CECAPGFRGTTTCQR-----ICSPGFYGHRC	624
Db	922	S--DGINMFFCNCBPAGFRGPKCEEDINECASNPCKNGANCTDCVNSYTCTCQPGFSGIHC	979
Qy	625	SQTCPPQCVHSS----GPCHHITGL----CDCLPGFTGALC----NE-----	658
Db	980	ESNTPDCTESSCFNGGTC--IDGINTFTCQCPPGFTGSGYQHDINECD SKPCLNGGTCQD	1037
Qy	659	-----VCPSGRFGKNCAGI-----CTCTNNGTC---NPIDRSCQCYPGWIGSDCSQP	702
Db	1038	SYGTYKCTCPQGYTG LNCQNLVRWCDSSPCKNGGKCWQTNNFYR-CECKSGWTGVYCDVP	1096
Qy	703	---CPPA--HWGPNCIHTCNCHNGAFC--SAYDGECKCTPGWTGLYCTQRCPLGFY GKDC	755
Db	1097	SVSCEVAAKQQGV DIVHL--CRNSGMCVDTGNTHFCRCQAGYTGSYCEEQV-----DEC	1148
Qy	756	ALICQCQNGADC-DHISG-QCTCRTGFMGRHCEQK-----	788
Db	1149	S-PNPCQNGATCTDYLGGSCECVAGYHGVCNSEEINECLSHPCQNGGTCIDLINTYKCS	1207
Qy	789	CPSGTGYGCRQICD-----CLNNSTC-DHITG-TCYCSPGWKGARCDQAG	832
Db	1208	CPRGTQGVHCEINVDDCTPFYDSFTLEPKCFNNGKCIDRVGGYNCICPPGFVGERCE---	1264
Qy	833	VIIIVGNLNS-LSRTSTALPADS	853
Db	1265	----GDVNECLSN-----PCDS	1277

RESULT 2

NTC1_HUMAN

ID NTC1_HUMAN STANDARD; PRT; 2556 AA.
AC P46531;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (hN1)
DE (Translocation-associated notch protein TAN-1).
GN NOTCH1 OR TAN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Mann R.S., Blaumueller C.M., Zagouras P.;
 RT "Complete human notch 1 (hN1) cDNA sequence.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE OF 1-2444 FROM N.A.
 RX MEDLINE=91347367; PubMed=1831692;
 RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
 RA Smith S.D., Sklar J.;
 RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
 RT chromosomal translocations in T lymphoblastic neoplasms.";
 RL Cell 66:649-661(1991).
 RN [3]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 RN [4]
 RP INTERACTION WITH DTX1.
 RX MEDLINE=98250176; PubMed=9590294;
 RA Matsuno K., Eastman D., Mitsiades T., Quinn A.M., Carcanciu M.L.,
 RA Ordentlich P., Kadesch T., Artavanis-Tsakonas S.;
 RT "Human deltex is a conserved regulator of Notch signalling.";
 RL Nat. Genet. 19:74-78(1998).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May be important for normal lymphocyte
 CC function. In altered form, may contribute to transformation or
 CC progression in some T-cell neoplasms. Involved in the maturation
 CC of both CD4+ and CD8+ cells in the thymus (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity). Interacts with DTX1 and DTX2.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -!- TISSUE SPECIFICITY: In fetal tissues most abundant in spleen,
 CC brain stem and lung. Also present in most adult tissues where it
 CC is found mainly in lymphoid tissues.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a

CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- DISEASE: NOTCH1 truncation is associated with T-cell acute
 CC lymphoblastic leukemia.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF308602; AAG33848.1; -.
 DR EMBL; M73980; AAA60614.1; -.
 DR HSSP; P00740; 1EDM.
 DR Genew; HGNC:7881; NOTCH1.
 DR MIM; 190198; -.
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0003793; F:defense/immunity protein activity; NAS.
 DR GO; GO:0006955; P:immune response; NAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 20.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 18.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2556 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.

FT	CHAIN	1722	2556	NOTCH EXTRACELLULAR TRUNCATION (BY
FT				SIMILARITY).
FT	CHAIN	1755	2556	NOTCH INTRACELLULAR DOMAIN (BY
FT				SIMILARITY).
FT	DOMAIN	19	1736	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1737	1757	POTENTIAL.
FT	DOMAIN	1758	2556	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	20	58	EGF-LIKE 1.
FT	DOMAIN	59	99	EGF-LIKE 2.
FT	DOMAIN	102	139	EGF-LIKE 3.
FT	DOMAIN	140	176	EGF-LIKE 4.
FT	DOMAIN	178	216	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	218	255	EGF-LIKE 6.
FT	DOMAIN	257	293	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	295	333	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	335	371	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	372	410	EGF-LIKE 10.
FT	DOMAIN	412	450	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	452	488	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	490	526	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	528	564	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	566	601	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	603	639	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	641	676	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	678	714	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	716	751	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	753	789	EGF-LIKE 20.
FT	DOMAIN	791	827	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	829	868	EGF-LIKE 22.
FT	DOMAIN	870	906	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	908	944	EGF-LIKE 24.
FT	DOMAIN	946	982	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	984	1020	EGF-LIKE 26.
FT	DOMAIN	1022	1058	EGF-LIKE 27.
FT	DOMAIN	1060	1096	EGF-LIKE 28.
FT	DOMAIN	1098	1144	EGF-LIKE 29.
FT	DOMAIN	1146	1182	EGF-LIKE 30.
FT	DOMAIN	1184	1220	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1222	1266	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1268	1306	EGF-LIKE 33.
FT	DOMAIN	1308	1347	EGF-LIKE 34.
FT	DOMAIN	1349	1385	EGF-LIKE 35.
FT	DOMAIN	1388	1427	EGF-LIKE 36.
FT	REPEAT	1446	1481	LIN/NOTCH 1.
FT	REPEAT	1482	1523	LIN/NOTCH 2.
FT	REPEAT	1524	1563	LIN/NOTCH 3.
FT	REPEAT	1928	1957	ANK 1.
FT	REPEAT	1961	1991	ANK 2.
FT	REPEAT	1995	2024	ANK 3.
FT	REPEAT	2028	2057	ANK 4.
FT	REPEAT	2061	2090	ANK 5.
FT	DOMAIN	1576	1579	POLY-VAL.
FT	DOMAIN	1662	1665	POLY-ARG.
FT	DOMAIN	1729	1732	POLY-PRO.
FT	DOMAIN	1741	1744	POLY-ALA.
FT	DOMAIN	1902	1905	POLY-GLU.
FT	DOMAIN	2260	2263	POLY-GLY.

FT	DOMAIN	2404	2407	POLY-GLN.
FT	DOMAIN	2411	2418	POLY-PRO.
FT	DOMAIN	2522	2525	POLY-SER.
FT	SITE	1665	1666	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY
FT				SIMILARITY).
FT	DISULFID	24	37	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.

Query Match 15.3%; Score 1034.5; DB 1; Length 2556;
Best Local Similarity 25.8%; Pred. No. 3.1e-54;
Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

Qy	94	CCPGFYESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGWTNCSSACDGDH	143
Db	89	CALGF--SGPLCLTPLDNACLTNPCRNGGTCDLLTLLEYKCRCPGWSGKSCQQA-----	141
Qy	144	WGPHTSRCQCKNGALCNPIITGA--CHCAAGFRGWRCE---DRCEQG----TYGNDCHQ-	193
Db	142	--DPCASN-PCANGGQCLPFEASYICHCPPSFHGPTCRQDVNECGQKPRLCRHGGTCHNE	198
Qy	194	-----RC-----QCQNGATC---DHVTGECRCPPGYTGAFCE---	222
Db	199	VGSYRCVCRATHTGPN CERPYVPCSPSPCQNGGTCRPTGDVTHECACLPGFTGQNC EENI	258
Qy	223	DLCPPG--KHGPQC-----EQRCP-----CQNGGVCHHVTG-	251
Db	259	DDCPGNCKNGGACVDGVNTYNCPCPPEWTGQYCTEDVDECQLMPNACQNGGTCHNTHGG	318
Qy	252	-ECSCPSGWMGTVCGQ-----PCPEGRFGKNC--SQEC--	281
Db	319	YNCVCVNGWTGEDCSENIDDCASAA CFHGATCHDRVASFYCECPHGRTGLLLCHLNDACIS	378
Qy	282	-QCHNGGTCD A--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLC AETCQCVNNGGKCY	333
Db	379	NPCNEGSNCDTNPVNGKAICTCPSGYTG PACSQDVDECS LGAN-----PCEHAGKCI	430
Qy	334	HVSGA--CLCEAGFAGERCEARLCPEGLYGIKCDKRC---PCHLENTHSCHPMSGE--CA	386
Db	431	NTLGSFECQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM	477
Qy	387	CKPGWSGLYC-----NE---TCSPGFYGEACQ---QICS---C	415
Db	478	CMPGYEGVHC EVNTDECASSPCLHN GRCLDKINEFQCECPTGFTGHL CQYDVDECASTPC	537
Qy	416	QNGADC-----DSV-TGKCTCAPGFKG	436

Db 538 KNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSKDGVATFTCLCRPGYTG 597
 Qy 437 IDCST-----PCPL-----GTYGINCS-----SRCGCKNDAVCS 465
 | | | | | | | | | | : | :
 Db 598 HHCETNINECSSQPCRLRGTCQDPDNAYLCFCLKGTTGPNCEINLDDCASSPCDSGTCLD 657
 Qy 466 PVDG-SCTCKAGWHGVDCSIR-----CPSGTWGFGCNL-----TC----- 499
 : | | | : | : | : | | | | | |
 Db 658 KIDGYECACEPGYTGSMCNSNIDECAGNPCHNGGTCEGNGFTCRCPEGYHDPCLSEV 717
 Qy 500 -----QCLNGGACNTLDG-TCTCAPGWRGEKCEL----- 527
 | : | : | : | | | | | : :
 Db 718 NECNSNPCVHGACRDSLNGYKCDGDPGWSGTNCDINNNECESNPCVNGGTCKDMTSGIVC 777
 Qy 528 PCQDGTGYGLNCAERCD-----CSHADGC-HPTTGH-CRCLPGWSGVHCDSDV---CAEG- 575
 | : | | | | : | : | | : | : | |
 Db 778 TCREGFGSPNCQTNINECASNPCLNKGTCIDDVAGYKCNCLLPYTGATCEVVLAPCAPSP 837
 Qy 576 -RWGPNC-----SLPCYC---KNGASCSPDDGICECAPGFRGTTCQRI-----CS 616
 | | | | | : | | : | : | |
 Db 838 CRNGGECRQSEDYESFSCVCPTAGAKGQTCQVDINECVLSPCRHGASCQNTHGXYRCHCQ 897
 Qy 617 PGFYGHRCSQTCPCQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
 | : | | | | : | | | | | | |
 Db 898 AGYSGRNCETDIDDC--RPNPCHNGGSCTDGINTAFCDCLPGFRGTFCCEEDINECASDPC 955
 Qy 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPG 693
 | : | | : | | | | | : | : | |
 Db 956 RNGANCTDCVDSYTCTCPAGFSGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPG 1013
 Qy 694 WIGSDC-----SQP-----CPPAHWGPNC---IHTCN---CHNGA 722
 : | | | | : | | : | | | : | : | |
 Db 1014 FTGSYCQHVVNECDSRPCLLGGTCQDGRGLHRCTCPQGYTGPNQCQNLVHWCDSSPCKNGG 1073
 Qy 723 FC----SAYDGECKCTPGWTGLYCTQ----- 744
 | : | | : | | | | |
 Db 1074 KCWQTHTQY--RCECPSGWTGLYCDVPSVSCEVAAQROQGVDDVARLCQHGGCLVDAGNTHH 1131
 Qy 745 -RCPLGFYKDCALI---CQ---CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 | | | : | | : | | | | | : | : | :
 Db 1132 CRCQAGYTGSYCEDLVDECSPPCQNGATCTDYLGYSCKCVAGYHGVNCSEIIDECLSH 1191
 Qy 789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG 815
 | | | | | | | | | | : |
 Db 1192 PCQNGGTCLDLPNTYKCS CPRGTQGVHCEINVDDCNPPVDPVSRSPKCFNNGTCVDQVGG 1251
 Qy 816 -TCYCSPGWKGARCDQAGVIVGNLN 840
 : | | | : | | : | : |
 Db 1252 YSCTCPPGFVGERCE-----GDVN 1270

RESULT 3

NTC1_MOUSE

ID NTC1_MOUSE STANDARD; PRT; 2531 AA.
 AC Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9R0X7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Motch A)
 DE (mT14) (p300).
 GN NOTCH1 OR MOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch.";
 RL Genomics 15:259-264(1993).
 RN [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RX MEDLINE=93050801; PubMed=1426644;
 RA Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
 RT "Expression analysis of a Notch homologue in the mouse embryo.";
 RL Dev. Biol. 154:377-387(1992).
 RN [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744(1992).
 RN [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE=99364499; PubMed=10437788;
 RA Lee J.S., Ishimoto A., Yanagawa S.I.;
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
 RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
 RL FEBS Lett. 455:276-280(1999).
 RN [6]
 RP SEQUENCE OF 1950-2201 FROM N.A.
 RX MEDLINE=98029496; PubMed=9384671;
 RA Messerle M., Follo M., Nehls M., Eggert H., Boehm T.;
 RT "Dynamic changes in gene expression during in vitro differentiation of
 RT mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
 RN [7]

RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
 RP MUTAGENESIS OF 1651-ARG--ARG-1654.
 RX MEDLINE=98318619; PubMed=9653148;
 RA Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G.,
 RA Israel A.;
 RT "The Notch1 receptor is cleaved constitutively by a furin-like
 RT convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
 RN [8]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [9]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 RN [10]
 RP INTERACTION WITH DTX1 AND DTX2.
 RX MEDLINE=21123790; PubMed=11226752;
 RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
 RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
 RA Okano H., Matsuno K.;
 RT "Murine homologs of deltex define a novel gene family involved in
 RT vertebrate Notch signaling and neurogenesis.";
 RL Int. J. Dev. Neurosci. 19:21-35(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation. May be involved in mesoderm
 CC development, somite formation and neurogenesis. Involved in the
 CC maturation of both CD4+ and CD8+ cells in the thymus.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds. Interacts with DTX1 and DTX2.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q01705-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
 CC thymus. Expressed at lower levels in the spleen, bone-marrow,
 CC spinal cord, eyes, mammary gland, liver, intestine, skeletal

CC muscle, kidney and heart.
 CC -!- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
 CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
 CC endothelial cells, while much lower levels are seen in the
 CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
 CC the neuroepithelium. At 13.5 dpc expressed in the surface
 CC ectoderm, eye and developing whisker follicles.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z11886; CAA77941.1; -.
 DR EMBL; L02613; AAK14898.1; -.
 DR EMBL; X68278; CAA48339.1; -.
 DR EMBL; AJ238029; CAB40733.1; -.
 DR EMBL; X82562; CAA57909.1; -.
 DR PIR; A46019; A46019.
 DR PIR; B49175; B49175.
 DR HSSP; P00740; 1EDM.
 DR MGD; MGI:97363; Notch1.
 DR GO; GO:0005887; C:integral to plasma membrane; IC.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0030154; P:cell differentiation; IMP.
 DR GO; GO:0007386; P:compartment specification; IMP.
 DR GO; GO:0007219; P:N signaling pathway; IC.
 DR GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 7.

```

DR      Pfam; PF00008; EGF; 35.
DR      Pfam; PF00066; notch; 3.
DR      PIRSF; PIRSF002279; Notch; 1.
DR      PRINTS; PR00010; EGFBLOOD.
DR      PRINTS; PR00011; EGFLAMININ.
DR      PRINTS; PR01452; NOTCH.
DR      SMART; SM00248; ANK; 6.
DR      SMART; SM00179; EGF_CA; 24.
DR      SMART; SM00004; NL; 2.
DR      PROSITE; PS50297; ANK_REP_REGION; 1.
DR      PROSITE; PS50088; ANK_REPEAT; 2.
DR      PROSITE; PS00010; ASX_HYDROXYL; 22.
DR      PROSITE; PS00022; EGF_1; 34.
DR      PROSITE; PS01186; EGF_2; 27.
DR      PROSITE; PS50026; EGF_3; 36.
DR      PROSITE; PS01187; EGF_CA; 21.
KW      Receptor; Transcription regulation; Activator; Differentiation;
KW      Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW      Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW      Alternative splicing.
FT      SIGNAL          1          18          POTENTIAL.
FT      CHAIN           19       2531      NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT      CHAIN          1711       2531      NOTCH EXTRACELLULAR TRUNCATION.
FT      CHAIN          1744       2531      NOTCH INTRACELLULAR DOMAIN.
FT      DOMAIN          19       1725      EXTRACELLULAR (POTENTIAL).

```

Query Match 15.2%; Score 1028; DB 1; Length 2531;
Best Local Similarity 25.7%; Pred. No. 7.6e-54;
Matches 314; Conservative 83; Mismatches 286; Indels 538; Gaps 73;

Qy	86	TMYYRRKSQCCPGFYESGEMCV---PHCADKCVH-GRCI---APNTCQCEPGWGGTNCSS-	137
Db	121	TLTEYKCRCSPGW--SGKSCQQADPCASNPCANGGQCLPFESSYICRPPGFHGPTRQD	178
Qy	138	-----ACDGDHWGPHC-----TSRCQCKNGALCNP----	162
Db	179	VNECSQNPGLCRHGGHCHNEIGSYRCACCATHTGPHCELPHYVPCSPSPCQNGATCRPTGD	238
Qy	163	ITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCNQNGATC-DHV-TGECRCPPGYTGAF	220
Db	239	TTHECACLPGFAGQNCEENVN-----DCPGN-NCKNGGACVDGVNTYNCRCPP EVTGQY	291
Qy	221	C-EDLCPGPKHGPQCEQRCRP--CQNGGVCHHVTG--ECSCPSGWMGTVCQG-----	266
Db	292	CTEDV-----DEC-QLMPNACQNAGTCHNTHGGYNCVCVNGWTGEDCSENIDDCASAA	343
Qy	267	-----PCPEGRFQKNC--SQEC---QCHNGGTCDA--ATGQ--CHCSPG	301
Db	344	CFQGATCHDRVASFYCECPHGRTGLLCHLKHACISNPCNEGSNCDTNPVNGKRICTCPSG	403
Qy	302	YTGERCQ---DECPVGTYGVLCAETCQCVNGGKCYHVSGA--CLCEAGFAGERCEARLCP	356
Db	404	YTGPACSQDVDECDLGAN-----RCEHAGKCLNTLGSFECQCLQGYTGPGCEIDV--	453
Qy	357	EGLYGIKCDKRC---PCHLENTHSCHPMSE--CACKPGWSGLYCNET-----	399
Db	454	-----NECISNPC--QNATCLDQIGEFQICMPGYEGVYCEINTDECASSPCLHN	502

Qy 400 -----CSPGFYGEACQ---QICS---CQNGADC--DSVTGKCTCAPGFKGID 438
 | | | | | : | : | | | | : |
 Db 503 GHCMDKIHEFQCQCPKGFNGHLCQYDVDECASTPCKNGAKCLDGPNTYTCVCTEGYTGTGTH 562
 Qy 439 CST-----PCPLGT-----YGINCSSRCG-----CKNDAVCSFVD 468
 | | | : | : | : | : | |
 Db 563 CEVDIDECDPDPCHYGSCKDGVATFTCLCQPGYTGHHCETNINECHSQPCRHHGGTCQDRD 622
 Qy 469 GS--CTCKAGWHGVDCSIR-----CPSGTWGFGCNLTQCCLNGGACNTLDG-TCTCA 517
 | | | | | : | | | | | : | | | |
 Db 623 NSYLCLCLKGTTGPNCEINLDDCASNPDCSGT-----CL-----DKIDGYECACE 667
 Qy 518 PGWRGEKCEL-----PCQDGTGYLNCARCDCSHADGCHPTT----- 554
 | : | | : | : | | | | : | | |
 Db 668 PGYTGSMLCNVNIDECAGSPCHNGGTCEDEGIAGFTC--RC----PEGYHDPCLSEVNECN 721
 Qy 555 -----GHCR-----CLPGWSGVHCD-----SVCAE 574
 | | | | | : | | | | |
 Db 722 SNPCIHGACRDGLNGYKCDCAPGWSGTNCDINNNECESNPCVNGGTCKDMTSGYVCTCRE 781
 Qy 575 GRWGPNC-----SLPCY-----CKNG 590
 | | | | | | | | | | |
 Db 782 GFSGPNCQTNINECASNPCLNQGTICIDDVAGYKCNCPYPYTGTATCEVVLAPCATSPCKNS 841
 Qy 591 ASCSPDDGI---CECAPGFRGTTTCQ-----RICSPGFYF 621
 | : | | | : | | | : | | : |
 Db 842 GVCKESEDYESFSCVCPTGWQGTCEVDINECVKSPCRHGASCQNTNGSYRCLCQAGYTG 901
 Qy 622 HRCSQTCPCQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
 | | | | | : | | | | | | | | | |
 Db 902 RNCESDIDDC--RPNPCHNGGSCTDGINTAFCDCCLPGFQGAFCCEEDINECASNPCQNGAN 959
 Qy 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPGWIGSD 698
 | | | | : | | | | | : | : | | : | |
 Db 960 CTDCVDSYTCPCPVGFNGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGFTGSY 1017
 Qy 699 C-----SQP-----CPPAHWGPNC---IHTCN---CHNGAFC--- 724
 | : | | : | | | : | : | | | |
 Db 1018 CQYDVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVRWCDSAPCKNGGRCWQT 1077
 Qy 725 -SAYDGECKCTPGWTGLYC---TQRCPLGFY--GKDCALICQ----- 760
 : | | : | | | : | : | | | : | |
 Db 1078 NTQY--HCECRSGWTGVNCDVLSVSCEVAAQKRGIDVTLLCQHGGLCVDEGDKHYCHCQA 1135
 Qy 761 -----CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 | | | | | : | | | | : | : | : |
 Db 1136 GYTGSYCEDEVDECSNPNCQNGATCTDYLGGFSCCKCVAGYHGSNCSEEINECLSQPCQNG 1195
 Qy 789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG-TCYC 819
 | | | | | | | | | | : | | | | |
 Db 1196 GTCIDLTNSYKCS CPRGTQGVHCEINVDDCHPPLDPASRSPKCFNNGTCVDQVGGYTCTC 1255
 Qy 820 SPGWKGARCDQAGVIIIVGNLN 840
 | : | | : | : |
 Db 1256 PPGFVGECERCE-----GDVN 1269

RESULT 4

NTC1_RAT

ID NTC1_RAT STANDARD; PRT; 2531 AA.
AC Q07008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
RT development.";
RL Development 113:199-205(1991).
RN [2]
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP FUNCTION.
RX MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT neural progenitor cells to an astroglial fate.";
RL Neuron 29:45-55(2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=93202015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene.";
RL Development 116:931-941(1992).
RN [5]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT functional roles for the Notch-DSL signaling system during brain
RT development.";
RL J. Comp. Neurol. 436:167-181(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte

CC suppression.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
 CC Expressed in postnatal central nervous system (CNS) germinal zones
 CC and, in early postnatal life, within numerous cells throughout the
 CC CNS. Found in both subventricular and ventricular germinal zones.
 CC -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
 CC days 12 and 14 and decrease rapidly to much lower levels in the
 CC adult.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X57405; CAA40667.1; -.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 25.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 35.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT CHAIN 1711 2531 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1724 1746 POTENTIAL.
 FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 58 EGF-LIKE 1.
 FT DOMAIN 59 99 EGF-LIKE 2.
 FT DOMAIN 102 139 EGF-LIKE 3.
 FT DOMAIN 140 176 EGF-LIKE 4.
 FT DOMAIN 178 216 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 218 255 EGF-LIKE 6.
 FT DOMAIN 257 293 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 295 333 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 335 371 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 372 410 EGF-LIKE 10.
 FT DOMAIN 412 450 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 452 488 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 490 526 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 528 564 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 566 601 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 603 639 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 641 676 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 678 714 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 716 751 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 753 789 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 829 867 EGF-LIKE 22.
 FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 907 943 EGF-LIKE 24.
 FT DOMAIN 945 981 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 983 1019 EGF-LIKE 26.
 FT DOMAIN 1021 1057 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1059 1095 EGF-LIKE 28.
 FT DOMAIN 1097 1143 EGF-LIKE 29.
 FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1267 1305 EGF-LIKE 33.
 FT DOMAIN 1307 1346 EGF-LIKE 34.
 FT DOMAIN 1348 1384 EGF-LIKE 35.

FT	DOMAIN	1387	1426	EGF-LIKE 36.
FT	REPEAT	1445	1480	LIN/NOTCH 1.
FT	REPEAT	1481	1522	LIN/NOTCH 2.
FT	REPEAT	1523	1562	LIN/NOTCH 3.
FT	REPEAT	1917	1946	ANK 1.
FT	REPEAT	1950	1980	ANK 2.
FT	REPEAT	1984	2013	ANK 3.
FT	REPEAT	2017	2046	ANK 4.
FT	REPEAT	2050	2079	ANK 5.
FT	DOMAIN	1730	1733	POLY-ALA.
FT	DOMAIN	1891	1894	POLY-GLU.
FT	DOMAIN	2258	2261	POLY-PRO.
FT	DOMAIN	2497	2500	POLY-SER.
FT	SITE	1654	1655	CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY
FT				SIMILARITY).
FT	DISULFID	24	37	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.
FT	DISULFID	206	215	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.

Query Match 15.2%; Score 1024; DB 1; Length 2531;
 Best Local Similarity 25.8%; Pred. No. 1.3e-53;
 Matches 315; Conservative 83; Mismatches 286; Indels 536; Gaps 74;

Qy	86	TMYRRKSQCCPGFYESGEMCV---PHCADKCVH-GRCI---APNTCQCEPGWGGTNCSS-	137
		: : : : : : : : : : : : :	
Db	121	TLTEYKCRCPFGW--SGKSCQQADPCASNPCANGGQCLPFESSYICGCPPGFHGPCTCRQD	178
Qy	138	-----ACDGDHWGPHC-----TSRCQCKNGALCNP----	162
Db	179	VNECSQNPLCRHGGTCHNEIGSYRCACRATHTGPHCELPHYVPCSPSPCQNGGTCRPTGD	238
Qy	163	ITGACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQNGATC-DHV-TGECRCPPGYTGAF	220
		:	
Db	239	TTHCACLPGFAGQNCEENV-----DCPGN-NCKNGGACVDGVNTYNCRCPPPEWTGQY	291
Qy	221	C-EDLCPPGKHGPQCEQRCP--CQNGGVCHHVTG--ECSCPSGWMGTVCQG-----	266
		:	
Db	292	CTEDV-----DEC-QLMPNACQNAQTCHNSHGGYNCVCVNGWTGEDCSDNIDDCASAA	343
Qy	267	-----PCPEGRFGKNC--SQEC---QCHNGGTCD--ATGQ--CHCSPG	301
		:	
Db	344	CFQGATCHDRVASFYCECPHGRTGLLCHLNDACISNPCNEGSNCNTNPVNGKAICTCPRG	403

Qy 302 YTGERCQ---DECPVGTYGVLCAETCQCVNNGKCYHVSGA--CLCEAGFAGERCEARLCP 356
 ||| | ||| :| | : ||| : | : | | | : | ||| :
 Db 404 YTGPACSQDVDECALGAN-----PCEHAGKCLNTLGSFECQCLQGYTGPRCEIDV-- 453

 Qy 357 EGLYGIKCDKRC---PCHLENTHSCHPMSGE--CAKPGWSGLYC----- 396
 | || :| :| || | | || :| :||
 Db 454 -----NECISNPC--QNDATCLDQIGEFQCICMPGYEGVYCEINTDECASSPCLHN 502

 Qy 397 -----NE---TCSPGFYGEACQ---QICS---CQNGADC--DSVTGKCTCAPGFKGID 438
 || | || | || | : | : || | | | | : |
 Db 503 GRCVDKINEFLCQCPKGFSGHLCQYDVDECASTPCKNGAKCLDGPNTYTCVCTEGYTGTH 562

 Qy 439 CST-----PCPLGTYGINCSSRCGCKNDAVCS PVDGSCTCKAGW----- 477
 | || :| || | | : | : || :|
 Db 563 CEVDIDECDPDPCHIGL-----CK-DGVAT---FTCLCQPGYTGHHCETNINECH 608

 Qy 478 -----HGVDCSIR-----CPSGTWGFGCNLTQ-----CLNGGACNTLDG-TCTCAP 518
 || | | | || | | : | : | : || | | |
 Db 609 SQPCRHHGGTCQDRDNYYLCLCLKGTTGPNCEINLDDCASNPCDSGTCLDKIDGYECACEP 668

 Qy 519 GWRGEKCEL-----PCQDGTYGLNCAERCDCSHADGCHPTT----- 554
 | : | | : | : || | | | | : | | |
 Db 669 GYTGSMCNVNIDECAGSPCHNGGTCELDGIAGFTC--RC----PEGYHDPTCLSEVNECN 722

 Qy 555 -----GHCR-----CLPGWSGVHCD-----SVCAEG 575
 | || | || || | : || | | |
 Db 723 NPCIHGACRDGLNGYKCDCAPGWSGTNCDINNNECESNPCVNGGTCKDMTSGYVCTCREG 782

 Qy 576 RWGPNC-----SLPCY-----CKNGA 591
 |||| | || ||||
 Db 783 FSGPNCQTNINECASNPCLNQGTICDDVAGYKCNCPYPYTGATCEVVLAPCATSPCKNSG 842

 Qy 592 SCSPDDGI---CECAPGFRGTTCQ-----RICSPGFYGH 622
 | : | | | : : || : | : |
 Db 843 VCKESEDYESFSCVCPTGWQGQTCEIDINECVKSPCRHGASCQNTNGSYRCLCQAGYTGR 902

 Qy 623 RCSQTCPCQCVHSSGPCHH-----ITGLCDCLPGFTGALCNE----- 658
 | | ||| : |||| || | |
 Db 903 NCESDIDDC--RPNPCHNGGSCTDGVNAAFCDCLPGFQGAFCCEEDINECATNPCQNGANC 960

 Qy 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCQCYPGWIGSDC 699
 || :| | :| || | | ||| :| :| | || :| |
 Db 961 TDCVDSYTCTCPTGFNGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGFTGSYC 1018

 Qy 700 -----SQP-----CPPAHWGPNC---IHTCN---CHNGAFC--- 724
 | :| || :| || | :| :| | || |
 Db 1019 QYDVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVRWCDSAPCKNGGKCWQTN 1078

 Qy 725 SAYDGECKCTPGWTGLYC---TQRCPLGFY--GKDCALICQ----- 760
 : | | :| |||| | : | : | | :||
 Db 1079 TQY--HCECRSGWTGFNCDVLSVSCEVAAQKRGIDVTLLCQHGGLCVDEEDKHYCHCQAG 1136

 Qy 761 -----CQNGADC-DHISG-QCTCRTGFMGRHCEQK----- 788
 |||| | | : : | | | : | : | :
 Db 1137 YTGSYCEDEVDECSPNPCQNGATCTDYLGGFSCCKCVAGYHGSNCSEEINECLSQPCQNGG 1196

 Qy 789 -----CPSGTYGYGCRQICD-----CLNNSTC-DHITG-TCYCS 820

				:
Db	1197	TCIDLTNTYKCS	PRGTQGVHCEIN	VDDCHPPLDPASRSPKCFNNGTCVDQVGGYTCTCP 1256
QY	821	PGWKGARCDQAG	VIIVGNNLN	840
		: :	::	
Db	1257	PGFVGERCE-----	GDVN	1269

RESULT 5

NTC2_HUMAN

ID NTC2_HUMAN STANDARD; PRT; 2471 AA.
AC Q04721; Q99734; Q9H240;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (hN2).
GN NOTCH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Blaumueller C.M., Mann R.S.;
RT "Complete human notch 2 (hN2) cDNA sequence.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.G.;
RT "Human Notch2, a novel member of cell-fate determining NOTCH
RT family.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 967-1229 FROM N.A.
RC TISSUE=T-cell;
RA Lemasson I., Devaux C., Mesnard J.M.;
RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE OF 1810-2447 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93265135; PubMed=1303260;
RA Stifani S., Blaumueller C.M., Redhead N.J., Hill R.E.,
RA Artavanis-Tsakonas S.;
RT "Human homologs of a Drosophila enhancer of split gene product define
RT a novel family of nuclear proteins.";
RL Nat. Genet. 2:119-127(1992).
RN [5]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=97386453; PubMed=9244302;
RA Blaumueller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on
RT the plasma membrane.";
RL Cell 90:281-291(1997).
RN [6]

RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor."
RL Am. J. Pathol. 154:785-794(1999).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity).
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung,
CC skeletal muscle and liver.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 35 EGF-like domains.
CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 6 ANK repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF308601; AAA36377.2; -.
DR EMBL; AF315356; AAG37073.1; -.
DR EMBL; U77493; AAB19224.1; -.
DR HSSP; P00740; 1EDM.
DR Genew; HGNC:7882; NOTCH2.
DR MIM; 600275; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 29.
 DR PROSITE; PS50026; EGF_3; 35.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2471 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1697 2471 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1678 1698 POTENTIAL.
 FT DOMAIN 1699 2471 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-LIKE 1.
 FT DOMAIN 64 102 EGF-LIKE 2.
 FT DOMAIN 105 143 EGF-LIKE 3.
 FT DOMAIN 144 180 EGF-LIKE 4.
 FT DOMAIN 182 219 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 221 258 EGF-LIKE 6.
 FT DOMAIN 260 296 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 298 336 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 338 374 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 375 413 EGF-LIKE 10.
 FT DOMAIN 415 454 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 456 492 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 494 530 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 532 568 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 570 605 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 607 643 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 645 680 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 682 718 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 720 755 EGF-LIKE 19.
 FT DOMAIN 757 793 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 795 831 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 833 871 EGF-LIKE 22.
 FT DOMAIN 873 909 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 911 947 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 949 985 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	987	1023	EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1025	1061	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1063	1099	EGF-LIKE 28.
FT	DOMAIN	1101	1147	EGF-LIKE 29.
FT	DOMAIN	1149	1185	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1187	1223	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1225	1262	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1264	1302	EGF-LIKE 33.
FT	DOMAIN	1304	1343	EGF-LIKE 34.
FT	DOMAIN	1374	1412	EGF-LIKE 35.
FT	REPEAT	1420	1456	LIN/NOTCH 1.
FT	REPEAT	1503	1535	LIN/NOTCH 2.
FT	REPEAT	1827	1871	ANK 1.
FT	REPEAT	1876	1905	ANK 2.
FT	REPEAT	1909	1939	ANK 3.
FT	REPEAT	1943	1972	ANK 4.
FT	REPEAT	1976	2005	ANK 5.
FT	REPEAT	2009	2038	ANK 6.
FT	DOMAIN	1645	1648	POLY-ALA.
FT	DOMAIN	1994	1997	POLY-LEU.
FT	DOMAIN	2426	2429	POLY-SER.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	79	BY SIMILARITY.
FT	DISULFID	73	90	BY SIMILARITY.
FT	DISULFID	92	101	BY SIMILARITY.
FT	DISULFID	109	121	BY SIMILARITY.
FT	DISULFID	115	131	BY SIMILARITY.
FT	DISULFID	133	142	BY SIMILARITY.
FT	DISULFID	148	159	BY SIMILARITY.
FT	DISULFID	153	168	BY SIMILARITY.
FT	DISULFID	170	179	BY SIMILARITY.
FT	DISULFID	186	198	BY SIMILARITY.
FT	DISULFID	192	207	BY SIMILARITY.
FT	DISULFID	209	218	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	230	246	BY SIMILARITY.
FT	DISULFID	248	257	BY SIMILARITY.

Query Match 15.0%; Score 1014.5; DB 1; Length 2471;
 Best Local Similarity 23.0%; Pred. No. 4.7e-53;
 Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

Qy	93	QCCPGFYESGEMCVPHCADKCVHGR-CIAPNTCQ-----CEPGWGGTNCSSACDG	141
		: : : : :	
Db	91	RCASGF--TGEDCQYSTSHPCFVSRPCLNGGTCHMLSRDYEECTCQVGFTGKEC-----	142
Qy	142	DHWGPHCTSRCQCKNGALCNPI TG--ACHCAAGFRGWRCEDRCEQGTYGNDCHQRCQCQN	199
		: : : : : :	
Db	143	-QWTDACLSH-PCANGSTCTTVANQFSCKCLTGFTGQKCE-----TDVNECDIPGHCHQ	194
Qy	200	GATCDHVTG--ECRCPPGYTGAFCEDL-----CPPGKHG	231
		: : : : :	
Db	195	GGTCLNLPGSYQCQCPCQGFTGQYCDSLYVPCAPSPCVNGGTCRQTGDFTFECNCLPGFEG	254
Qy	232	PQCEQ-----RCP-----CQNGGV	245

Db	255	STCERNIDDCPNHRCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVDECLLQPNACQNGGT	314
Qy	246	CHHVTG--ECSCPSGWMGTVCQGQ-----PCPEGRFGKNC-	277
Db	315	CANRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASFSCMCPGKAGLLCH	374
Qy	278	-SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQC	326
Db	375	LDDACISNPCHKGALCDTNPLNGQYICTCPQGYKGADCTEDVDECAM-----ANSNPC	427
Qy	327	VNGGKCYHVSQA--CLCEAGFAGERCE-----ARLCPEGLY	360
Db	428	EHAGKCVNTDGAHFCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFK	487
Qy	361	GIKCDKR-----CP-----CHLE-----NTHSC	378
Db	488	GVHCELEINECQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQIDIDDCSSTPCLNGAKC	547
Qy	379	--HPMSGECACKPGWSGLYCNET-----CSPGFYGEAC-QQ	411
Db	548	IDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGDIDSYTCICNPGYMGALCSDQ	607
Qy	412	I-----CSCQNGA-----DCDS---VTG-----K	427
Db	608	IDECYSSPCLNDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCASNPICHGICMDGINRYS	667
Qy	428	CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGCK-----NDAVCS	465
Db	668	CVCSPGFTGQRCNIDIDECASNPCRKGATCINGVNGFRICPEGPHHPSCYSQVNECLSN	727
Qy	466	P-VDGSCT-----CKAGWHGVDCSI-----RCPSGT	490
Db	728	PCIHGNCTGGLSGYKCLDAGWVGINCEVDKNECLSNPCQNGGTCDNLVNGYRCTCKKGF	787
Qy	491	WGFGCNLTQ-----CLNGGAC-----	507
Db	788	KGYNCQVNIDECASNPCNLQGTCTFDDISGYTCHCVLPYTGKNCQTVLAPCSPNPCENAAV	847
Qy	508	-----NTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGH--CRCL	560
Db	848	CKESPNEFESYTCLCAPGWQQRCTIDIDE-----CISK-PCMNHGLCHNTQGSYMCECP	900
Qy	561	PGWSGVHCDVCAEGRWGPNCSLPCYCKNGASCSPDDGI----CECAPGFRGTTCQR---	613
Db	901	PGFSGMDCEEDIDDCLANP-----CQNGGSCM--DGVNTFSCLCPLPGFTGDKCQTDMN	951
Qy	614	-----ICSPGFYGHRSQTCPCQCVHSS-----GPCHHITGL---	644
Db	952	ECLSEPCKNGGTCSDYVNSYTCKCQAGFDGVHCENNINECTESSCFNGGTG--VDGINSF	1009
Qy	645	-CDCLPGFTGALC----NEV-----CPSGRFGKNC---AGICT-	674
Db	1010	SCLCPVGFTGSFCLHEINECSSHPCLNEGTCVDGLGTYRCSCPLGYTGKNCQTLVNLCSR	1069
Qy	675	--CTNNGTC--NPIDRSCQCYPGWIGSDCSQP-----	702

RX MEDLINE=21331789; PubMed=11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May play an essential role in postimplantation
 CC development, probably in some aspect of cell specification and/or
 CC differentiation (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
 CC plexus in the brain. Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within numerous
 CC cells throughout the CNS. It is more highly localized to
 CC ventricular germinal zones. Also found in the heart, liver and
 CC kidney.
 CC -!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 35 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M93661; AAK13558.1; -.
 DR PIR; A49128; A49128.
 DR HSSP; P00743; 1CCF.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.

DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 35.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2471 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1697 2471 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1678 1698 POTENTIAL.
 FT DOMAIN 1699 2471 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-LIKE 1.
 FT DOMAIN 64 102 EGF-LIKE 2.
 FT DOMAIN 105 143 EGF-LIKE 3.
 FT DOMAIN 144 180 EGF-LIKE 4.
 FT DOMAIN 182 219 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 221 258 EGF-LIKE 6.
 FT DOMAIN 260 296 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 298 336 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 338 374 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 375 413 EGF-LIKE 10.
 FT DOMAIN 415 454 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 456 492 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 494 530 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 532 568 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 570 605 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 607 643 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 645 680 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 682 718 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 720 755 EGF-LIKE 19.
 FT DOMAIN 757 793 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 795 831 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	833	871	EGF-LIKE 22.
FT	DOMAIN	873	909	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	911	947	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	949	985	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	987	1023	EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1025	1061	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1063	1099	EGF-LIKE 28.
FT	DOMAIN	1101	1147	EGF-LIKE 29.
FT	DOMAIN	1149	1185	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1187	1223	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1225	1262	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1264	1302	EGF-LIKE 33.
FT	DOMAIN	1304	1343	EGF-LIKE 34.
FT	DOMAIN	1374	1412	EGF-LIKE 35.
FT	DOMAIN	1645	1648	POLY-ALA.
FT	DOMAIN	1994	1997	POLY-LEU.
FT	DOMAIN	2426	2429	POLY-SER.
FT	DOMAIN	2446	2451	POLY-GLY.
FT	REPEAT	1420	1456	LIN/NOTCH 1.
FT	REPEAT	1503	1535	LIN/NOTCH 2.
FT	REPEAT	1827	1871	ANK 1.
FT	REPEAT	1876	1905	ANK 2.
FT	REPEAT	1909	1939	ANK 3.
FT	REPEAT	1943	1972	ANK 4.
FT	REPEAT	1976	2005	ANK 5.
FT	REPEAT	2009	2038	ANK 6.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	79	BY SIMILARITY.
FT	DISULFID	73	90	BY SIMILARITY.
FT	DISULFID	92	101	BY SIMILARITY.
FT	DISULFID	109	121	BY SIMILARITY.
FT	DISULFID	115	131	BY SIMILARITY.
FT	DISULFID	133	142	BY SIMILARITY.
FT	DISULFID	148	159	BY SIMILARITY.
FT	DISULFID	153	168	BY SIMILARITY.
FT	DISULFID	170	179	BY SIMILARITY.
FT	DISULFID	186	198	BY SIMILARITY.
FT	DISULFID	192	207	BY SIMILARITY.
FT	DISULFID	209	218	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	230	246	BY SIMILARITY.
FT	DISULFID	248	257	BY SIMILARITY.
FT	DISULFID	264	275	BY SIMILARITY.
FT	DISULFID	269	284	BY SIMILARITY.
FT	DISULFID	286	295	BY SIMILARITY.
FT	DISULFID	302	315	BY SIMILARITY.
FT	DISULFID	309	324	BY SIMILARITY.
FT	DISULFID	326	335	BY SIMILARITY.
FT	DISULFID	342	353	BY SIMILARITY.
FT	DISULFID	347	362	BY SIMILARITY.
FT	DISULFID	364	373	BY SIMILARITY.
FT	DISULFID	379	390	BY SIMILARITY.
FT	DISULFID	384	401	BY SIMILARITY.
FT	DISULFID	403	412	BY SIMILARITY.
FT	DISULFID	419	433	BY SIMILARITY.

FT	DISULFID	427	442	BY SIMILARITY.
FT	DISULFID	444	453	BY SIMILARITY.
FT	DISULFID	460	471	BY SIMILARITY.
FT	DISULFID	465	480	BY SIMILARITY.
FT	DISULFID	482	491	BY SIMILARITY.
FT	DISULFID	498	509	BY SIMILARITY.
FT	DISULFID	503	518	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	536	547	BY SIMILARITY.

Query Match 14.8%; Score 998; DB 1; Length 2471;
 Best Local Similarity 24.4%; Pred. No. 4.6e-52;
 Matches 321; Conservative 79; Mismatches 322; Indels 596; Gaps 70;

Qy	17	CHWIGTASPLNLEDPNVCSHWESYSVTVQESY-----PHPFQIYYTSCTDIL	64
		: : : : : :	
Db	121	CHMLS-----WDTYECTCQVGFTGKQCQWTDVCLSHPCEN--GSTCSSVA	163
Qy	65	NWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKCVHGRCIAPNTC	124
		: : : :	
Db	164	NQFSC-----RCPAGI--TGQKCDADINECDIPGRCQHGGTC	198
Qy	125	QCEPGWGGTNCSSACDGDHWGPHCTS-----RCQCKNGALC---NPITGACHCAAGFRG	175
Db	199	LNLP--GSYRCQ--CPQRFTHGQCDSPYVPCAPSPCVNNGGTCRQTGDFSTSECHCLPGFEG	254
Qy	176	WRCEDRCEQGTYGNDCHQRCQCONGATC-DHV-TGECRCPPGYTGAFD-EDLCPPGKHGP	232
		: :	
Db	255	SNCERNID-----DCPNH-KCQNGGVCDGVNTYNCRCPPQWTGQFCTEDV-----D	300
Qy	233	QC-EQRCPCQNGGVCHHVTG--ECSCPSGWMGTVCQGP-----	267
		: : :	
Db	301	ECLLQPNACQNGGTCTNRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASF	360
Qy	268	---CPEGRFGKNC--SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCQ---DECP	312
		:	
Db	361	SCLCPEGKAGLLCHLDDACISNPCHKGALCDTNPLNGQYICTCPQAYKGADCTEDVDECA	420
Qy	313	VGTYGVLCAETCQCVNNGGKCYHVSQA--CLCEAGFAGERCE-----	351
		: : : : :	
Db	421	M-----ANSNPCEHAGKCVNTDGAHFCECLKGYAGPRCEMDINECHSDPCQNDATCLD	473
Qy	352	-----ARLCPEGLYGIKCDKR-----CP-----CHLE-	373
		: : : :	
Db	474	KIGGFTCLCMPGFGKVHCELEVNECQSNPCVNNGQCVDKVNRFOCLCPPGFTGPVCQIDI	533
Qy	374	-----NTHSC--HPMSGECACKPGWSGLYCNET-----	399
		: : :	
Db	534	DDCSSTPCLNGAKCIDHPNGYECQCATGFTGTLCDENIDNCDPDPCHHGQCQDGIDSYTC	593
Qy	400	-CSPGFYGEAC-QQICSC-----QNGADCDSVTG-KCTCAPGFKGIDC-----STP	442
		: : : : : :	
Db	594	ICNPGYMGAIQSDQIDECYSSPCLNDGRCIDLNVGYQCNCQPGTSGLNCEINFDDCASNP	653
Qy	443	CPLGTY--GIN-----CS-----SRCG-----CKNDAVC-----	464
		:	
Db	654	CLHGACVDGINRYSCVCSPGFTGQRCNIDIDECASNPCRKDATCINDVNGFRMCPEGPH	713

Qy 465 -----SP-VDGSCT-----CKAGWHGVDCSI----- 484
 || : ||: || | || | : || :
 Db 714 HPSCYSQVNECLSSPCIHGNCCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGGTCN 773

Qy 485 -----RCPSTWGFEGCNLTQ-----CLNGGAC----- 507
 | | | : | : | || | |
 Db 774 NLVNGYRCTCKKGFKGYNCQVNIDECASNPCLNQGTCLDDVSGYTCHCMLPYTGKNCQTV 833

Qy 508 -----NTLDGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHAD 548
 | | | || || || : || : | : | :
 Db 834 LAPCSPNPCENAAVCKEAPNFESFTCLCAPGWQGRCTVDVDE-----CVSK-PCMNG 886

Qy 549 GCHPTTGH--CRCLPGWGSVHCDVCAEGRWGPNCSLPCYCKNGASC--SPDDGICECAP 604
 || | | | | || : || : || : | | | : || || : | | |
 Db 887 ICHNTQGSYMCECPPGFSGMDCEE-----DINDCLANPCQNGGSCVDKVNTFSCCLCLP 939

Qy 605 GFRGTTQR-----ICSPGFYGHRSQTCPQCVHSS---- 635
 || | | | | | | | | : | | : | | |
 Db 940 GFVGDKCQTDMECLSEPCKNGGTCSYVNSYTCTCPAGFHGVHCENNIDECTESSCFNG 999

Qy 636 GPCHHITGL----CDCLPGFTGALC----NE-----VCPSGRFG 666
 | | : | : | | | | | | | | | | | |
 Db 1000 GTC--VDGINSFSCCLCPVGFTGPFCLHDINECSSNPCLNSGTCVDGLGTYRCTCPLGYTG 1057

Qy 667 KNC---AGICT---CTNNGTC--NPIDRSCQCYPGWIGSDCSQ----- 701
 || | : | : | | | | | | | | : |
 Db 1058 KNCQTLVNLCSFSPCKNKGTCQAEKARPRCLCPPGWDGAYCDVLNVSCKAAALQKGVPE 1117

Qy 702 -----PCPPAHWGPNC---IHTC---NCHNGAFCSAYDG--ECKCTP 735
 | | : | | : | | : || || : | | : | |
 Db 1118 HLCQHSGICINAGNTHHCQCPGLYTGSYCEEQLDEECASNPCQHGCATCSDFIGGYRCECVP 1177

Qy 736 GWTGLYCTQR-----CPLGFYG-----KDCALICQCQN 763
 | : | : | | | | | | | | | | | |
 Db 1178 GYQGVNCEYEVDECQNQPCQNGGTCIDLNVNHFKCSCPPGTRGLLCEENIDDCAGAPHCLN 1237

Qy 764 GADC-DHISG-----QCTCRTGFM 781
 | | | | | | | | : |
 Db 1238 GGQCVDRIGGYSCRCLPGFAGERCEGDINECLSNPCSSSEGLDCIQLKNNYQCVCERSAFT 1297

Qy 782 GRHCE---QKCPSTYGYGCRQICDCLNNSTCDHITGT-----CYCSPGWKGARCDQA 831
 || || | | | | | | : | | | : || || :
 Db 1298 GRHCETFLDVCPQK-----PCLNGGTCAVASNVDPGFICRCPGFSGARCS 1345

RESULT 7

NTC2_MOUSE

ID NTC2_MOUSE STANDARD; PRT; 2470 AA.
 AC O35516; Q06008; Q60941;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Motch
 DE B).
 GN NOTCH2.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and mutliiform transcripts encoded by a
 RT single copy of mouse Notch2 gene.";
 RL Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE OF 316-1518 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.
 RX MEDLINE=97075110; PubMed=8917536;
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
 RA Martin D.I.;
 RT "Inhibition of granulocytic differentiation by mNotch1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=99396706; PubMed=10393120;
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 RT embryonic lethality.";
 RL Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RX MEDLINE=95333893; PubMed=7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 RT mouse brain.";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular

CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O35516-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O35516-2; Sequence=VSP_001405;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
 CC neuroepithelia, somites, optic vesicles and branchial arches, but
 CC not heart.
 CC -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
 CC the postnatal ependymal cells, and the choroid plexus throughout
 CC embryonic and postnatal development.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 35 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D32210; BAA22094.1; -.
 DR EMBL; X68279; CAA48340.1; -.
 DR EMBL; U31881; AAC52924.1; -.
 DR PIR; A49175; A49175.
 DR HSSP; P16109; 1FSB.
 DR MGD; MGI:97364; Notch2.
 DR GO; GO:0005887; C:integral to plasma membrane; IC.
 DR GO; GO:0005515; F:protein binding; IPI.

DR GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
 DR GO; GO:0007219; P:N signaling pathway; IC.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 2.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS50026; EGF_3; 35.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2470 NOTCH EXTRACELLULAR TRUNCATION.
 FT CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1678 1698 POTENTIAL.
 FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-LIKE 1.
 FT DOMAIN 64 102 EGF-LIKE 2.
 FT DOMAIN 105 143 EGF-LIKE 3.
 FT DOMAIN 144 180 EGF-LIKE 4.
 FT DOMAIN 182 219 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).
 FT DOMAIN 258 294 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 296 334 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 336 372 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 373 411 EGF-LIKE 10.
 FT DOMAIN 413 452 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 454 490 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 492 528 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 530 566 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 568 603 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 605 641 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 643 678 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 680 716 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	718	753	EGF-LIKE 19.
FT	DOMAIN	755	791	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	793	829	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	831	869	EGF-LIKE 22.
FT	DOMAIN	871	907	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	909	945	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	947	983	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	985	1021	EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1023	1059	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1061	1097	EGF-LIKE 28.
FT	DOMAIN	1099	1145	EGF-LIKE 29.
FT	DOMAIN	1147	1183	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1185	1221	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1223	1260	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1262	1300	EGF-LIKE 33.
FT	DOMAIN	1302	1345	EGF-LIKE 34.
FT	DOMAIN	1372	1410	EGF-LIKE 35.
FT	REPEAT	1418	1454	LIN/NOTCH 1.
FT	REPEAT	1501	1533	LIN/NOTCH 2.
FT	REPEAT	1825	1869	ANK 1.

Query Match 14.7%; Score 993; DB 1; Length 2470;
 Best Local Similarity 25.8%; Pred. No. 9.1e-52;
 Matches 316; Conservative 87; Mismatches 295; Indels 526; Gaps 71;

Qy	93	QCCPGFYESGEMCVPHCADKCVHGR-CIAPNTCQ-----CEPGWGGTNCSSACDG	141
		: : : : :	
Db	91	RCAPGF--TGEDCQYSTSHPCFVSRPCQNGGTCHMLSRDTYECTCQVGFTGKQC-----	142
Qy	142	DHWGPHCTSRCQCKNGALCNPITG--ACHCAAGFRGWRCE-----DRCEQ-----	184
		: : : : : :	
Db	143	-QWTDACLSH-PCENGSTCTSVASQFSCKCPAGLTGQKCEADINECDIPGRCQHGGTCLN	200
Qy	185	--GTYGNDG-----HQRGQ-----CQNGATCDHV--TGECRCPPGYTGAFCE---D	223
		:	
Db	201	LPGSYRCQCGQFTGQHCDSPYVRGLPCVNGGTCRQTGDFTECNCLPGFEGSTCERNID	260
Qy	224	LCPPGK--HGPQCEQ-----RCP-----CQNGGVCHHVTVG--	251
		: :	
Db	261	DCPNHKCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVDECLLQPNACQNGGTCTNRNGGY	320
Qy	252	ECSCPSGWMGTVCQP-----CPEGRFGKNC--SQEC---	281
		: :	
Db	321	GCVCVNGWSGDDCSENIDDCAYASCTPGSTCIDRVASFSCLCPEGKAGLLCHLDDACISN	380
Qy	282	QCHNGGTCDA--ATGQ--CHCSPGYTGERCQ--DECPVGTYGVLCAETCQCVNNGGKCYH	334
		: : : :	
Db	381	PCHKGALCDTNPLNGQYICTCPQGYKGADCTEDVDECAM-----ANSNPCEHAGKCVN	433
Qy	335	VSGA--CLCEAGFAGERCE-----ARLCPEGLYGIKCDKR-	367
		:	
Db	434	TDGAHFHCECLKGYAGPRCEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFGKVHCELEV	493
Qy	368	-----CP-----CHLE-----NTHSC--HPMSG	384
		:	
Db	494	NECQSNPCVNNGQCVDKVNRQCLCPPGFTGPVCQIDIDDCSSTPCLNGAKCIDHPNGYE	553

Qy	385	CACKPGWSGLYCNET-----CSPGFYGEAC-QQI-----	412
		: : : : : : : :	
Db	554	CQCATGFTGILCDENIDNCDPDPCHHGQCQDGIDSYTICNPGYMGAIACSDQIDECYSSP	613
Qy	413	-----CSCQNGA-----DCDS-----VTG----KCTCAPGFK	435
		: : :	
Db	614	CLNDGRCIDLVNGYQCNCQPGTSGLNCEINFDDCASNPCHMGVCVDGINRYSCVCSPGFT	673
Qy	436	GIDC-----STPCPLGTYGINCSS--RCGCK-----NDAVCSP-VDGSCT	472
		: : : : :	
Db	674	GQRCNIDIDECASNPCRKGATCINDVNGFRICPEGPHHPSCYSQVNECLSNPCIHGNT	733
Qy	473	-----CKAGWHGVDCSI-----RCPSGTWGFGCNLT	498
		: : : :	
Db	734	GGLSGYKCLCDAGWVGVNCEVDKNECLSNPCQNGGTCNNLVNGYRCTCKKGFKGYNCQVN	793
Qy	499	CQ-----CLNGGAC-----NTL	510
Db	794	IDECASNPCLNQGTCTFDDVSGYTCHCMLPYTGKNCQTVLAPCSPNPCENAAVCKEAPNFE	853
Qy	511	DGTCTCAPGWRGEKCELPQDGTYGLNCAERCDCSHADGCHPTTGH--CRCLPGWSGVHC	568
		: : : : : : : :	
Db	854	SFSCLCAPGWQGRCTVDVDE-----CISK-PCMNGVCHNTQGSYVCECPPPGFSGMDC	906
Qy	569	DSVCAEGRWGPNCSLPCYCKNGASCSPDDGI---CECAPGFRGTTTCQR-----	613
		: : : :	
Db	907	EE-----DINDCLANPCQNGGSCV--DHVNTFSCQCHPGFIGDKCQTDMNELSEPCCK	957
Qy	614	-----ICSPGFYGHRCSTCPQCVHSS---GPCHHITGL---CDCLPGF	651
		: : : :	
Db	958	NGGTCSDYVNSYTCTCPAGFHGVHCENNIDECTESSCFNGGTC--VDGINSFSCCLCPVGF	1015
Qy	652	TGALC---NE-----VCPSGRFGKNC---AGICT---CTNNGT	680
		: : : :	
Db	1016	TGPFCLHDINECSSNPCLNAGTCVDGLGTYRCICPLGYTGKNCQTLVNLCSRSPCKNKG	1075
Qy	681	C--NPIDRSCQCYPGWIGSDC---SQPCPPA--HWGPNCIHTCNCHNGAFCSAYD-GECK	732
		: : : : :	
Db	1076	CVQEKARPHCLCPPGWDGAYCDVLNVSCAAALQKGVPEHLCQ-HSGICINAGNTHHCQ	1134
Qy	733	CTPGWTGLYCTQRCPLGFYGKDCALICQCQNGADC-DHISG-QCTCRTGFMGRHCEQK---	788
		: : : : : : :	
Db	1135	CPLGYTGSYCEEQL-----DECA-SNPCQHGATCNDFIGGYRCECVPGYQGVNCEYEVD	1187
Qy	789	-----CPSGTYGYGCRQICD-----CLNNSTC-DHITG-T	816
		:	
Db	1188	ECQNQPCQNGGTCIDLNVNHFKSCPPGTRGLLCEENIDECAGGPHCLNGGQCVDRIGGYT	1247
Qy	817	CYCSPGWKGARCDQAGVIIIVGNLN	840
		: : :	
Db	1248	CRCLPGFAGERCE-----GDIN	1264

RESULT 8

NTC1 BRARE

ID NTC1 BRARE STANDARD; PRT; 2437 AA.

AC P46530;

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor.
 GN NOTCH1A OR NOTCH.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=94128602; PubMed=8297791;
 RA Bierkamp C., Campos-Ortega J.A.;
 RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
 RT its pattern of transcription during early embryogenesis.";
 RL Mech. Dev. 43:87-100(1993).
 CC -!- FUNCTION: Implicated in cell fate specifications during
 CC embryo development. May be involved in the formation of the
 CC neural plate, notochord and brain vesicles.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
 CC stages. During gastrulation is differentially expressed,
 CC accumulating predominantly in the prechordal mesoderm and
 CC notochord. At the end of gastrulation, expressed along the
 CC anterior-posterior axis including the developing neural plate
 CC and differentiating mesoderm. Also present in the developing
 CC brain and head regions.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X69088; CAA48831.1; -.
 DR PIR; S42612; S42612.
 DR HSSP; P00740; 1EDM.
 DR ZFIN; ZDB-GENE-990415-173; notch1a.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.

DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 22.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 28.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Neurogenesis; Repeat; ANK repeat;
 KW EGF-like domain; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1725 1747 POTENTIAL.
 FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 21 57 EGF-LIKE 1.
 FT DOMAIN 58 98 EGF-LIKE 2.
 FT DOMAIN 101 138 EGF-LIKE 3.
 FT DOMAIN 139 175 EGF-LIKE 4.
 FT DOMAIN 177 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 217 254 EGF-LIKE 6.
 FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 294 332 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 334 370 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 371 409 EGF-LIKE 10.
 FT DOMAIN 411 449 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 451 487 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 489 524 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 526 562 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 564 599 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 601 637 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 639 674 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 676 712 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 714 749 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 751 787 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 789 825 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 827 865 EGF-LIKE 22.
 FT DOMAIN 867 903 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 905 941 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 943 979 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 981 1017 EGF-LIKE 26.
 FT DOMAIN 1019 1055 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1057 1093 EGF-LIKE 28.
 FT DOMAIN 1095 1141 EGF-LIKE 29.
 FT DOMAIN 1143 1179 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1181 1217 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1219 1263 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1265 1303 EGF-LIKE 33.
 FT DOMAIN 1305 1344 EGF-LIKE 34.

FT	DOMAIN	1346	1382	EGF-LIKE 35.
FT	DOMAIN	1385	1423	EGF-LIKE 36.
FT	REPEAT	1446	1486	LIN/NOTCH 1.
FT	REPEAT	1487	1520	LIN/NOTCH 2.
FT	REPEAT	1521	1561	LIN/NOTCH 3.
FT	REPEAT	1867	1910	ANK 1.
FT	REPEAT	1915	1944	ANK 2.
FT	REPEAT	1948	1978	ANK 3.
FT	REPEAT	1982	2011	ANK 4.
FT	REPEAT	2015	2044	ANK 5.
FT	REPEAT	2048	2077	ANK 6.
FT	DOMAIN	2265	2276	POLY-GLN (OPA-REPEAT).
FT	SITE	1656	1657	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY
FT				SIMILARITY).
FT	DISULFID	25	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	73	BY SIMILARITY.
FT	DISULFID	67	86	BY SIMILARITY.
FT	DISULFID	88	97	BY SIMILARITY.
FT	DISULFID	105	116	BY SIMILARITY.
FT	DISULFID	110	126	BY SIMILARITY.
FT	DISULFID	128	137	BY SIMILARITY.
FT	DISULFID	143	154	BY SIMILARITY.
FT	DISULFID	148	163	BY SIMILARITY.
FT	DISULFID	165	174	BY SIMILARITY.
FT	DISULFID	181	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	503	BY SIMILARITY.
FT	DISULFID	498	512	BY SIMILARITY.
FT	DISULFID	514	523	BY SIMILARITY.
FT	DISULFID	530	541	BY SIMILARITY.
FT	DISULFID	535	550	BY SIMILARITY.
FT	DISULFID	552	561	BY SIMILARITY.
FT	DISULFID	568	578	BY SIMILARITY.

FT	DISULFID	573	587	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	605	616	BY SIMILARITY.
FT	DISULFID	610	625	BY SIMILARITY.
FT	DISULFID	627	636	BY SIMILARITY.
FT	DISULFID	643	653	BY SIMILARITY.
FT	DISULFID	648	662	BY SIMILARITY.
FT	DISULFID	664	673	BY SIMILARITY.
FT	DISULFID	680	691	BY SIMILARITY.
FT	DISULFID	685	700	BY SIMILARITY.
FT	DISULFID	702	711	BY SIMILARITY.
FT	DISULFID	718	728	BY SIMILARITY.
FT	DISULFID	723	737	BY SIMILARITY.
FT	DISULFID	739	748	BY SIMILARITY.
FT	DISULFID	755	766	BY SIMILARITY.
FT	DISULFID	760	775	BY SIMILARITY.
FT	DISULFID	777	786	BY SIMILARITY.
FT	DISULFID	793	804	BY SIMILARITY.
FT	DISULFID	798	813	BY SIMILARITY.
FT	DISULFID	815	824	BY SIMILARITY.
FT	DISULFID	831	842	BY SIMILARITY.
FT	DISULFID	836	853	BY SIMILARITY.
FT	DISULFID	855	864	BY SIMILARITY.
FT	DISULFID	871	882	BY SIMILARITY.
FT	DISULFID	876	891	BY SIMILARITY.
FT	DISULFID	893	902	BY SIMILARITY.
FT	DISULFID	909	920	BY SIMILARITY.

Query Match 14.6%; Score 987; DB 1; Length 2437;
 Best Local Similarity 24.8%; Pred. No. 2.1e-51;
 Matches 310; Conservative 81; Mismatches 320; Indels 538; Gaps 70;

Qy	91	KSQCCPGFYESGEMCVPHCADKCVHGRC-----IAPNTCQCEPGWGGTNCSSACD	140
		: : :	
Db	85	KCDCVLGF--SDRLCLTPVNHACMNSPCRNGGTCSLTLDTFTCRCQPGWSGKTCQLA--	140
Qy	141	GDHWGPHCTSRCQCKNGALCNPITG--ACHCAAGFRGWRCEDRCEQGTYNDCCH--QRCQC	197
		:	
Db	141	-----DPCASN-PCANGGQCSAFESHYICTCPPNFHGQTCRQDV-----NECAVSPSPC	188
Qy	198	QNGATCDHVTGE--CRCPPGYTGAFCEDL-----CPPGK	229
		: :	
Db	189	RNGGTCINEVGSYLCRCPPPYTGPHCQRLYQPCLPSPCRSGGTCVQTSDDTHTCCLPGF	248
Qy	230	HGPQCE-----QRC-----PCQNG	243
Db	249	TGQTCEHNVDDCTQHACENGGPICIDGINTYNCHCDKHWGTGQYCTEDVDECELSPNACQNG	308
Qy	244	GVCHHVTG--ECSCPSGWMGTVCQ-----PCPEGRFGKN	276
		: :	
Db	309	GTCHNTIGGFHCVCVNGWTGDDCSENIDDCASAAACSHGATCHDRVASFFCECPHGRTGLL	368
Qy	277	C--SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCQ---DECPVGTGYVLCAETC	324
		:	
Db	369	CHLDDACISNPCQKGSNCDTNPVSGKAICTCPPGYTGSACNQDIDECSLGAN-----	420
Qy	325	QCVNGGKCYHVSGA--CLCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELENTHSCHPMS	382

Db	421	PCEHGGRCCLNTKGSFQCKCLQGYEGPRCEMDV-----NEC-KSNPC--QNDATCLDQI	470
Qy	383	G--ECACKPGWWSGLYCNET-----CSPGFYGEACQ---QIC	413
Db	471	GGFHCICMPGYEGVFCQINSDDCASQPCLNGKCIDKINSFHCECPKGFSGSLCQVDVDEC	530
Qy	414	S---CQNGADCDSVTGK--CTCAPGFKGIDC-----STPCPL-----GTYGINCSSR	455
Db	531	ASTPCKNGAKCTDGPKNYTCECTPGFSGIHCELDINECASSPCHYGVCRDGVASFTCDCCR	590
Qy	456	CG-----CKNDAVCSPVDGS--CTKAGWHGVDCSIR-----	485
Db	591	PGYTGRLCETNINECLSQPCRNNGGTCQDRENAYICTCPKGTGTVNCEINIDDCKRKPCDY	650
Qy	486	-----CPSGTWGFGCNLTCQ-----CLNGGACNTLDG---TCTCAPGWRG	522
Db	651	GKCIDKINGYECVCEPGYSGSMCNINIDDCALNPCHNGGTC--IDGVNSFTCLCPDGFRD	708
Qy	523	EKC-----ELPCQDGTYGLNC---AERC---DCSHADGCHP	552
Db	709	ATCLSQHNECSSNPCIHGSCLDQINSYRCVCEAGWMGRNCDININECLSNPCVNGGTCKD	768
Qy	553	-TTGH-CRCLPGWSGVHCDSVCAEGRWGP-----NCSL-----	583
Db	769	MTSGYLCTCRAGFSGPNCQMNINECASNPCLNQGSCIDDVAGFKCNCMLPYTGEVCENVL	828
Qy	584	-PCY---CKNGASCSPDD-----GICE-----	601
Db	829	APCSPRPCKNGGVCRESEDFQSFSCNCPAGWQGQTCEVDINECVRNPCTNGGVCCENLRGG	888
Qy	602	----CAPGFRGTTTCQR-----ICSPGFYGHRCSTCPQCV	632
Db	889	FQCRCNPGFTGALCENDIDDCEPNPCSNGGVCQDRVNGFVCVCLAGFRGERCAEDIDECV	948
Qy	633	HSSGPCHHITGLCDCLPGFTGALCNEVCPSGRFGKNC---AGICT---CTNNGTCNPIDR	686
Db	949	--SAPCRNNGGNCTDCVNSYT---CS--CPAGFSGINCEINTPDCTESSCFNGGTC--VDG	999
Qy	687	----SCQCYPGWIGSDC-----SQP-----CPPAHWGPNC---IH	714
Db	1000	ISSFSCVCLPGFTGNYCQHDVNECDSRPCQNGGSCQDGYGTYKCTCPHGYTGLNCQSLVR	1059
Qy	715	TCN---CHNGAFC--SAYDGECKCTPGWTGLYCTQ-----	744
Db	1060	WCDSSPCKNGGSCWQQGASFTQCASGWTGIYCDVPSVSCEVAARQQGVSVAVLCRHAGQ	1119
Qy	745	-----RCPLGFGYKDC---ALICQ---CQNGADC-DHISG-QCTCRTGFMGRHCE	786
Db	1120	CVDAGNTHLCRCQAGYTGSYCQEQQVDECQPNPCQNGATCTDYLGGYSCECVPGYHGMNCS	1179
Qy	787	QK-----CPSGTYGYGCRQICD-----CLNN	807
Db	1180	KEINECLSQPCQNGGTCIDLVTYKCS CPRGTQGVHCEIDIDDCSPSVDPILTGEPRCFNG	1239
Qy	808	STC-DHITG-TCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTALPADSY	854

RESULT 9

NOTC_DROME

ID NOTC_DROME STANDARD; PRT; 2703 AA.
AC P07207; O97458; P04154; Q9W4T8;
DT 01-NOV-1986 (Rel. 03, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurogenic locus Notch protein precursor.
GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=86079539; PubMed=3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats.";
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S, and Oregon-R; TISSUE=Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of Drosophila melanogaster: relationship
RT of the encoded protein to mammalian clotting and growth factors.";
RL Mol. Cell. Biol. 6:3094-3108(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
 RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of *D.*
 RT *melanogaster*.";
 RL Science 287:2220-2222(2000).
 RN [5]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE=85099329; PubMed=2981631;
 RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
 RT "opa: a novel family of transcribed repeats shared by the Notch locus
 RT and other developmentally regulated loci in *D. melanogaster*.";
 RL Cell 40:55-62(1985).
 RN [6]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=87257846; PubMed=3037327;
 RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
 RT "Restriction of P-element insertions at the Notch locus of *Drosophila*
 RT *melanogaster*.";
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RN [7]
 RP INTERACTION WITH DX, AND MUTANT SU42C.
 RX MEDLINE=94215489; PubMed=8162848;
 RA Diederich R.J., Matsuno K., Hing H., Artavanis-Tsakonas S.;
 RT "Cytosolic interaction between *deltex* and Notch ankyrin repeats
 RT implicates *deltex* in the Notch signaling pathway.";
 RL Development 120:473-481(1994).

RN [8]
 RP INTERACTION WITH DX.
 RX MEDLINE=95401878; PubMed=7671825;
 RA Matsuno K., Diederich R.J., Go M.J., Blaumueller C.M.,
 RA Artavanis-Tsakonas S.;
 RT "Deltex acts as a positive regulator of Notch signaling through
 RT interactions with the Notch ankyrin repeats.";
 RL Development 121:2633-2644(1995).
 RN [9]
 RP S3 CLEAVAGE BY PSN.
 RX MEDLINE=99221487; PubMed=10206646;
 RA Struhl G., Greenwald I.;
 RT "Presenilin is required for activity and nuclear access of Notch in
 RT Drosophila.";
 RL Nature 398:522-525(1999).
 RN [10]
 RP S3 CLEAVAGE BY PSN.
 RX MEDLINE=99221488; PubMed=10206647;
 RA Ye Y., Lukinova N., Fortini M.E.;
 RT "Neurogenic phenotypes and altered Notch processing in Drosophila
 RT Presenilin mutants.";
 RL Nature 398:525-529(1999).
 RN [11]
 RP S2 CLEAVAGE BY KUZ.
 RX MEDLINE=21657146; PubMed=11799064;
 RA Lieber T., Kidd S., Young M.W.;
 RT "kuzbanian-mediated cleavage of Drosophila Notch.";
 RL Genes Dev. 16:209-221(2002).
 RN [12]
 RP MUTANT MCD5.
 RX MEDLINE=21575956; PubMed=11719214;
 RA Ramain P., Khechumian K., Seugnet L., Arbogast N., Ackermann C.,
 RA Heitzler P.;
 RT "Novel Notch alleles reveal a Deltex-dependent pathway repressing
 RT neural fate.";
 RL Curr. Biol. 11:1729-1738(2001).
 RN [13]
 RP REVIEW.
 RX MEDLINE=22256570; PubMed=12369105;
 RA Portin P.;
 RT "General outlines of the molecular genetics of the Notch signalling
 RT pathway in Drosophila melanogaster: a review.";
 RL Hereditas 136:89-96(2002).
 CC -!- FUNCTION: Signaling protein, which regulates, with both positive
 CC and negative signals, the differentiation of at least central and
 CC peripheral nervous system and eye, wing disk, oogenesis, segmental
 CC appendages such as antennae and legs, and muscles, through lateral
 CC inhibition or induction. Functions as a receptor for membrane-
 CC bound ligands Delta and Serrate to regulate cell-fate
 CC determination. Upon ligand activation, and releasing from the cell
 CC membrane, the Notch intracellular domain (NICD) forms a
 CC transcriptional activator complex with Su(H) (Suppressor of
 CC hairless) and activates genes of the E(spl) complex. Essential for
 CC proper differentiation of ectoderm.
 CC -!- SUBUNIT: Interacts with Su(H) when activated. Interacts with Dx
 CC via its ANK repeats.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Upon activation and

```

CC      S3 cleavage, it is released from the cell membrane and enters into
CC      the nucleus in conjunction with Su(H).
CC      -!- PTM: Upon binding its ligands such as Delta or Serrate, it is
CC      cleaved (S2 cleavage) in its extracellular domain, close to the
CC      transmembrane domain. S2 cleavage is probably mediated by Kuz. It
CC      is then cleaved (S3 cleavage) downstream of its transmembrane
CC      domain, releasing it from the cell membrane. S3 cleavage requires
CC      Psn.
CC      -!- SIMILARITY: Belongs to the NOTCH family.
CC      -!- SIMILARITY: Contains 36 EGF-like domains.
CC      -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC      -!- SIMILARITY: Contains 6 ANK repeats.

```

```

DR      EMBL; M16152; AAB59220.1; -.
DR      EMBL; M16153; AAB59220.1; JOINED.
DR      EMBL; M16149; AAB59220.1; JOINED.
DR      EMBL; M16150; AAB59220.1; JOINED.
DR      EMBL; M16151; AAB59220.1; JOINED.
DR      EMBL; K03508; AAA28725.1; -.
DR      EMBL; M13689; AAA28725.1; JOINED.
DR      EMBL; K03507; AAA28725.1; JOINED.
DR      EMBL; AE003426; AAF45848.2; -.
DR      EMBL; AL035436; CAB37610.1; -.
DR      EMBL; AL035395; CAB37610.1; JOINED.
DR      EMBL; M12175; AAA74496.1; -.
DR      EMBL; M16025; AAA28726.1; -.

```

Query Match 14.5%; Score 978.5; DB 1; Length 2703;
Best Local Similarity 26.8%; Pred. No. 7.4e-51;
Matches 290; Conservative 102; Mismatches 297; Indels 395; Gaps 70;

Qy	7	SCL---- <td>60</td>	60
		: : : : :::: : :	
Db	502	SCLDDPGTFRVCVCM PGFTGTQCEIDIDECQSNPC-----LNDGTC	541
Qy	61	TDILNWEFKCTRHRVSYRTAYRHGEKTM YRRKSQCCPGFYESGEMC---VPHCADKCVHGR	117
		: : : : :	
Db	542	HDKINGFKCS-----CALGF--TGARCQINIDDCQSQPCRNR	576
Qy	118	-----CIAPNTCQCEPGWGGTNCS---SACDGDHWGPHCTSRCQCKNGALCNPITG-ACH	168
		: : : : : : : :	
Db	577	GICHDSIAGYSCECPPGYTGTSC EININDCDSN-----PCHRGKCIDDVNSFKCL	626
Qy	169	CAAGFRGW-----RCEDR-----CEQGTYG-----NDCHQRCQ	196
		: : : : :	
Db	627	CDPGYTG YICQKQINECESNPCQFDGHCQDRVGSYYCQCQAGTSGKNCEVNVNECHSN-P	685
Qy	197	CQNGATC-DHVTG-ECRCPPGYTGAFCEDLCP PGKHGPQCEQRCPCQNGGVC-HHVTG-E	252
		: : : : : : :	

Db 686 CNNGATCIDGINSYKQCQVPGFTGQHCE-----KNVDECIS-SPCANNGVCIDQVNGYK 738
 Qy 253 CSCPSGWMGTVC-----GQP-----CPEGRFGKNCS---QECQ-- 282
 | | | | : | | | | | | | | | |
 Db 739 CECPRGFYDAHCLSDVDECASNPCVNEGRCEDGINEFICHCPPGYTGKRCELDIDECSSN 798
 Qy 283 -CHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNGGKCY-HV 335
 | : | | | | | | | | | | : | : | | | | | |
 Db 799 PCQHGGTCYDKLNAFSCQCMPPGYTGQKCETNIDDC-----VTNFCGNGGTCTIDKV 848
 Qy 336 SG-ACLCEAGFAGERCEARLCPEGLYGIKC-DKRCPCHELENTHSCHPMSG---ECACKP 389
 : | | : | : | | | : : | | | : | | | | | |
 Db 849 NGYKCVCKVPFTGRDCESKMDP-----CASNRC---KNEAKCTPSSNFLDFSCCTCKL 897
 Qy 390 GWSGLYCNETCSPGFYGEACQQICSCQNGADCDSVTG--KCTCAPGFKGIDC----- 439
 | : | | | : | : | | | | : | | : | | | | |
 Db 898 GYTGRYCDEDI-----DECSLSSPCRNGASCLNVPGSYRCLCTKGYEGRDCAINTDDCA 951
 Qy 440 STPCP-----LGTYGINC-----SSRC-----GCKNDAVCSPVDGS--CTCK 474
 | | | : | | | | | | | | | | | | | |
 Db 952 SFPCQNGGTCLDGIGDYSCLCVDGFDGKHCETDINECLSQPCQNGATCSQYVNSYTCTCP 1011
 Qy 475 AGWHGVDCSIR-----CPSGTWGFGCN-----LTCQ-----CLN 503
 | : | : | | | | : | : | | | | | | | |
 Db 1012 LGFSGINCQTNDEDCTESSCLNGGSCIDGINGYNCSCLAGYSGANCQYKLNKCDSNPCLN 1071
 Qy 504 GGACNTLDG--TCTCAPGWRGEKC-----ELPCQDGTYGLNCAERCDCSHADGCHPT 553
 | | : : | | | | : | : | : | | | | | |
 Db 1072 GATCHEQNNEYTCHCPSGFTGKQCSEYVDWCGQSPCENG-----ATCSQMK--HQF 1120
 Qy 554 TGHCRCLPGWSGVHCD---SVCAEGRWGPNCSLPCYCKNGASCSP--DDGICECAPGFRG 608
 : | : | | : | | | : | | | | : | : | : | : |
 Db 1121 S--CKCSAGWTGKLCQDAADRKGLSLRQLCNNG-TCKDYGNSHVCYCSQGYAG 1177
 Qy 609 TTCQR-----ICSPGFYGHRCSTCPQCV---HSSGPCH 639
 : | : | | | | | | | | : | | | | |
 Db 1178 SYCQKEIDECQSQPCQNGGTCTDLIGAYECQCRQGFQGNCELNIDDCAPNPCQNGGTCH 1237
 Qy 640 H--ITGLCDCLPGFTGALC---NEVCPSGRFGKNCAIGICTCTNNGTCNPIDR---SCQC 690
 : | | | | : | : | | | | | | | | | |
 Db 1238 DRVMNFSCSCPPGTMGIICEINKDDCKPG-----ACHNNGSC--IDRVGGFECVC 1285
 Qy 691 YPGWIGSDC-----SQPCP-----PAHWGPNCIHTCN----- 717
 | | : | : | | | | | | | | : | | | : |
 Db 1286 QPGFVGARCEGDINECLSNPCSNAGTLDCVQLVNNYHCNCRPGHMGRHCEHKVDFCAQSP 1345
 Qy 718 CHNGAFCSAYDGECKCTPGWTGLYCTQRCPLGFYKGKDCALICQCCQNGADCDHIS---GQC 774
 | | | | : | : | | | | | : | | | | | |
 Db 1346 CQNGGNCNIRQ-----SGHHCI--CNNGFYGKNCEL-----SGQDCDSNPCRVGNC 1389
 Qy 775 TCRTGFMGRHCEQKCPSTGYGCR---QICD---CLNNSTCDHITG--TCYCSPGWK 825
 | | | | | | | | | | : | : | | | | | |
 Db 1390 VVADEGFGYRCE--CPRGTLGEHCEIDTLDECSPNPCAQGAACEDLLGDYECLCPSKWKG 1447
 Qy 826 ARCD 829
 | | |
 Db 1448 KRCD 1451

RESULT 10

NTC3_MOUSE

ID NTC3_MOUSE STANDARD; PRT; 2318 AA.
AC Q61982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NOTCH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR X Swiss Webster;
RX MEDLINE=95001556; PubMed=7918097;
RA Lardelli M., Dalstrand J., Lendahl U.;
RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT growth factor-repeats and is expressed in proliferating
RT neuroepithelium.";
RL Mech. Dev. 46:123-136(1994).
RN [2]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [3]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play a role during CNS
CC development.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Proliferating neuroepithelium.
CC -!- DEVELOPMENTAL STAGE: CNS development.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-

CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 34 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74760; CAA52776.1; -.
 DR PIR; S45306; S45306.
 DR HSSP; P00740; 1EDM.
 DR MGD; MGI:99460; Notch3.
 DR GO; GO:0005887; C:integral to plasma membrane; IC.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007219; P:N signaling pathway; IC.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 33.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 19.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS50026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_CA; 16.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.

FT	SIGNAL	1	39	POTENTIAL.
FT	CHAIN	40	2318	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT	CHAIN	1630	2318	NOTCH EXTRACELLULAR TRUNCATION.
FT	CHAIN	1663	2318	NOTCH INTRACELLULAR DOMAIN.
FT	DOMAIN	40	1643	EXTRACELLULAR.
FT	TRANSMEM	1644	1664	POTENTIAL.
FT	DOMAIN	1665	2318	CYTOPLASMIC.
FT	DOMAIN	2242	2261	PEST.
FT	DOMAIN	40	78	EGF-LIKE 1.
FT	DOMAIN	79	119	EGF-LIKE 2.
FT	DOMAIN	120	157	EGF-LIKE 3.
FT	DOMAIN	159	196	EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	198	235	EGF-LIKE 5.
FT	DOMAIN	237	273	EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	275	313	EGF-LIKE 7.
FT	DOMAIN	315	351	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	352	390	EGF-LIKE 9.
FT	DOMAIN	392	430	EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	432	468	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	470	506	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	508	544	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	546	581	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	583	619	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	621	656	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	658	694	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	696	731	EGF-LIKE 18.
FT	DOMAIN	735	771	EGF-LIKE 19.
FT	DOMAIN	772	809	EGF-LIKE 20.
FT	DOMAIN	811	848	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	850	886	EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	888	923	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	925	961	EGF-LIKE 24.
FT	DOMAIN	963	999	EGF-LIKE 25.
FT	DOMAIN	1001	1035	EGF-LIKE 26.
FT	DOMAIN	1037	1083	EGF-LIKE 27.
FT	DOMAIN	1085	1121	EGF-LIKE 28.
FT	DOMAIN	1123	1159	EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1161	1204	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1206	1245	EGF-LIKE 31.
FT	DOMAIN	1247	1288	EGF-LIKE 32.
FT	DOMAIN	1290	1326	EGF-LIKE 33.
FT	DOMAIN	1336	1374	EGF-LIKE 34.
FT	REPEAT	1388	1428	LIN/NOTCH 1.
FT	REPEAT	1429	1467	LIN/NOTCH 2.
FT	REPEAT	1468	1503	LIN/NOTCH 3.
FT	REPEAT	1839	1868	ANK 1.
FT	REPEAT	1872	1902	ANK 2.
FT	REPEAT	1906	1935	ANK 3.
FT	REPEAT	1939	1968	ANK 4.
FT	REPEAT	1972	2001	ANK 5.
FT	SITE	1572	1573	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).
FT				
FT	DISULFID	43	55	BY SIMILARITY.
FT	DISULFID	49	66	BY SIMILARITY.
FT	DISULFID	68	77	BY SIMILARITY.
FT	DISULFID	83	94	BY SIMILARITY.
FT	DISULFID	88	107	BY SIMILARITY.

FT	DISULFID	109	118	BY SIMILARITY.
FT	DISULFID	124	135	BY SIMILARITY.
FT	DISULFID	129	145	BY SIMILARITY.
FT	DISULFID	147	156	BY SIMILARITY.
FT	DISULFID	163	175	BY SIMILARITY.
FT	DISULFID	169	184	BY SIMILARITY.
FT	DISULFID	186	195	BY SIMILARITY.
FT	DISULFID	202	213	BY SIMILARITY.
FT	DISULFID	207	223	BY SIMILARITY.
FT	DISULFID	225	234	BY SIMILARITY.
FT	DISULFID	241	252	BY SIMILARITY.
FT	DISULFID	246	261	BY SIMILARITY.
FT	DISULFID	263	272	BY SIMILARITY.
FT	DISULFID	279	292	BY SIMILARITY.
FT	DISULFID	286	301	BY SIMILARITY.
FT	DISULFID	303	312	BY SIMILARITY.
FT	DISULFID	319	330	BY SIMILARITY.
FT	DISULFID	324	339	BY SIMILARITY.
FT	DISULFID	341	350	BY SIMILARITY.
FT	DISULFID	356	367	BY SIMILARITY.
FT	DISULFID	361	378	BY SIMILARITY.
FT	DISULFID	380	389	BY SIMILARITY.
FT	DISULFID	396	409	BY SIMILARITY.
FT	DISULFID	403	418	BY SIMILARITY.
FT	DISULFID	420	429	BY SIMILARITY.
FT	DISULFID	436	447	BY SIMILARITY.
FT	DISULFID	441	456	BY SIMILARITY.
FT	DISULFID	458	467	BY SIMILARITY.
FT	DISULFID	474	485	BY SIMILARITY.
FT	DISULFID	479	494	BY SIMILARITY.
FT	DISULFID	496	505	BY SIMILARITY.
FT	DISULFID	512	523	BY SIMILARITY.
FT	DISULFID	517	532	BY SIMILARITY.

Query Match 14.5%; Score 977.5; DB 1; Length 2318;
 Best Local Similarity 25.8%; Pred. No. 7.2e-51;
 Matches 322; Conservative 77; Mismatches 317; Indels 532; Gaps 74;

Qy	9	LSFICLLCHWIGTASPLNLEDPNVCSHWESYSVTVQESYPHPFDQIYYTSCTDILNWEK	68
		: :	
Db	62	LEAACLCPLPGWVG--ERCQLEDP--C-----HSGPCAGRGVCQSSVVAGTARFS	106
Qy	69	CTRHRVSYRTAYRHGEKTMYYRKSQCCPGFYESGEMCV---PHCADKCVHGR--CIAPN-	122
		: : : :	
Db	107	C-----RCLRGF--QGPDCSQPDPCVSRPCVHGAPCSVGPDPG	141
Qy	123	--TCQCEPGWGGTNCSSACDGDHWGPHC-----TSRCQ-----	153
		: : :	
Db	142	RFACACPPGYQGQSCQSDIDECRSGTTCRHGGTCLNTPGSFRCQCPLGYTGLLCENPVVP	201
Qy	154	-----CKNGALC---NPITGACHCAAGFRGWRCEQGGTYGNDCHQRCCQNGATC-D	204
		: : : : :	
Db	202	CAPSPCRNGGTCRQSSDVTYDCACLPGFEGQNCENVND-----DCPGH-RCLNGGTCVD	254
Qy	205	HV-TGECRCPPGYTGAF-C-EDLCPPGKHGPQCE-QRCPCQNGGVCHHVTG--ECSCPSGW	259
		: : : : :	
Db	255	GVNTYNCQCPPEWTGQFCTEDV-----DECQLQPNACHNGGTFCFNLLGGHSCVCVNGW	307

Qy 260 MGTVCQ-----PCPEGRFGKNC--SQEC---QCHNGGTC 289
 | | | | | : | | | | |
 Db 308 TGESCSQNIDDCATAVCFHGATCHDRVASFYCACPMGKTGLLCHLDDACVSNPCHEDAIC 367
 Qy 290 DA--ATGQ--CHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNNGKCYHVSGA--CL 340
 | : : | | : | | : | | : : | : : | : |
 Db 368 DTNPVSGRAICTCPPGFTGGACDQDVDECSIG-----ANPCEHL--GRCVNTQGSFLCQ 419
 Qy 341 CEAGFAGERCEARL--CPEGLYGIKCDKRCPCHLENTHSCHPMSGE--CACKPGWSGLYC 396
 | : | | | : | | | | : | | : | | : | |
 Db 420 CGRGYTGP RCETDVNECLSG-----PC--RNQATCLDRIGQFTCICMAGFTGTYC 467
 Qy 397 NE-----TCSPGFYGEACQ---QICS---CQNGADC-DSV 424
 | | | | | : | : | | |
 Db 468 EVDIDECQSSPCVNGGVCKDRVNGFSCTCPSGFSGSMCQLDVDECASTPCRNGAKCVDQP 527
 Qy 425 TG-KCTCAPGFKGI-----DCS-TPCPLGTYGINCSSRCGCKNDAVCS PVDG---SC 471
 | : | | | : | | | | | | | | | |
 Db 528 DGYECRCAEGFEGTLCERNVDDCSPDPCHHG-----RC-----VDGIASFSC 569
 Qy 472 TCKAGWHGVDCS-----IRCPSGTWGFGCNLTC-QCLNG- 504
 | : : | | | | | : | :
 Db 570 ACAPGYTGIRCESQVDECRSQPCRYGGKCLDLVDKYLRCPPGTTGVNCEVNIDDCASN 629
 Qy 505 ---GACNTLDG---TCTCAPGWRGEKCEL-----PCQDGTYGLNC----- 538
 | | | | | : | | : | | | : |
 Db 630 CTFGVCR--DGINRYDCVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFHCLCPPG 687
 Qy 539 -----AERCDCSHADGCHPTTG--HCRCLPGWSGVHCDSDVCAEGRWGPNC SLP 584
 | | | | | | | | | | | :
 Db 688 SLPPLCLPANHPCAHKPCSHG-VCHDAPGGFRCVCEPGWSGPRCSQSLA-----PDACES 741
 Qy 585 CYCKNGASCSPDDGI---CECAPGFRGTTCQRI-----CS 616
 | : | : : | | | | : | : : |
 Db 742 QPCQAGGTCT-SDGIGFRCTCAPGFGHQCEVLSPCTPSLCEHGGHCESDPDRLTVCSCP 800
 Qy 617 PGFYGHRCSTCPQCVHSS-----GPCHHITG--LCDCLPGFTGALCNE----- 658
 | : | | | : | : | | : | | : | :
 Db 801 PGWQGPRCQQDVDECA GASPCGPHGTCTNLPGNFRICHRGYTGPFCDQDIDDCDPNPCL 860
 Qy 659 -----VCPSGRFGKNCA-----GICT----- 674
 | | | | | | |
 Db 861 HGGSCQDGVGSFSCSCLDGFAGPRCARDVDECLSSPCGPGTCTDHVASFTCACPPGYGGF 920
 Qy 675 -----CTNNGTCNPIDR----SCQCYPGWIGSDC-----SQP-----C 703
 | | | | : | | | : | | : | : |
 Db 921 HCEIDL PDCSPSSCFNGGTC--VDGVSSFSCLCRPGYTGT HCQYEADPCFSRPCLHGGIC 978
 Qy 704 PPAHWGPNCIHTCN-----CHNGAFCSAYDGECKCTPGWTG----- 739
 | | | | | | | | | : |
 Db 979 NPTHPGFEC--TCREGFTGSQCQNPVDWCSQAPCQNGGRCVQTGAYCICPPGWSGRLCDI 1036
 Qy 740 --LYCTQR-----CPLGFYKDC-----ALICQCQN 763
 | | : | | | | | :
 Db 1037 QSLPCTEAAAQMGVRLEQLCQEGGKCIDKGRSHYCVCP EGRTGSHCEHEVD PCTAQPCQH 1096

Qy 764 GADCDHISG--QCTCRTGFMGRHCEQ-----KCPSTGYGY 796
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1097 GGTCTRGYMGGYVCECPAGYAGDSCEDNIDECASQPCQNGGSCIDLVARVLCSCPPGTLGV 1156

Qy 797 GC---RQICD-----CLNNSTCDHITG--TCYCSPGWKGARCD 829
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1157 LCEINEDDCDLGPSLDGSGVQCLHNGTCVDLVGGFRCNCPPGYTGLHCE 1204

RESULT 11

NTC3_HUMAN

ID NTC3_HUMAN STANDARD; PRT; 2321 AA.
 AC Q9UM47; Q9UEB3; Q9UPL3; Q9Y6L8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97032728; PubMed=8878478;
 RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
 RA Alamowitch S., Domenga V., Cecillion M., Marechal E., Maciazek J.,
 RA Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
 RA Bach J.-F., Bousser M.-G., Tournier-Lasserre E.;
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 RT causing stroke and dementia.";
 RL Nature 383:707-710(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gunel M., Artavanis-Tsakonas S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
 RA Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,
 RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
 RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 RA Carrano A.V.;
 RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
 RT 19p13.1.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
 RP ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
 RP GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
 RP CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
 RP ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
 RX MEDLINE=98049753; PubMed=9388399;
 RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,

RA Vayssiere C., Cruaud C., Maciazek J., Weissenbach J., Bousser M.-G.,
 RA Bach J.-F., Tournier-Lasserre E.;
 RT "Strong clustering and stereotyped nature of Notch3 mutations in
 RT CADASIL patients.";
 RL Lancet 350:1511-1515(1997).
 RN [5]
 RP VARIANT CADASIL 114-GLY--PRO-120 DEL.
 RX MEDLINE=20264473; PubMed=10802807;
 RA Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
 RA Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.;
 RT "Splice site mutation causing a seven amino acid Notch3 in-frame
 RT deletion in CADASIL.";
 RL Neurology 54:1874-1875(2000).
 RN [6]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
 CC tissues.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
 CC dominant arteriopathy with subcortical infarcts and
 CC leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type
 CC of stroke and dementia of which key features include recurrent
 CC subcortical ischemic events and vascular dementia.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 34 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U97669; AAB91371.1; -.
DR EMBL; AF058900; AAC14346.1; -.
DR EMBL; AF058881; AAC14346.1; JOINED.
DR EMBL; AF058882; AAC14346.1; JOINED.
DR EMBL; AF058883; AAC14346.1; JOINED.
DR EMBL; AF058884; AAC14346.1; JOINED.
DR EMBL; AF058885; AAC14346.1; JOINED.
DR EMBL; AF058886; AAC14346.1; JOINED.
DR EMBL; AF058887; AAC14346.1; JOINED.
DR EMBL; AF058888; AAC14346.1; JOINED.
DR EMBL; AF058889; AAC14346.1; JOINED.
DR EMBL; AF058890; AAC14346.1; JOINED.
DR EMBL; AF058891; AAC14346.1; JOINED.
DR EMBL; AF058892; AAC14346.1; JOINED.
DR EMBL; AF058893; AAC14346.1; JOINED.
DR EMBL; AF058894; AAC14346.1; JOINED.
DR EMBL; AF058895; AAC14346.1; JOINED.
DR EMBL; AF058896; AAC14346.1; JOINED.
DR EMBL; AF058897; AAC14346.1; JOINED.
DR EMBL; AF058898; AAC14346.1; JOINED.
DR EMBL; AF058899; AAC14346.1; JOINED.
DR EMBL; AC004257; AAC04897.1; -.
DR EMBL; AC004663; AAC15789.1; ALT_INIT.
DR PIR; S78549; S78549.
DR HSSP; P00740; 1EDM.
DR Genew; HGNC:7883; NOTCH3.
DR MIM; 600276; -.
DR MIM; 125310; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBLOOD.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 19.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.

DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 25.
 DR PROSITE; PS50026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_CA; 16.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Disease mutation.
 FT SIGNAL 1 39 POTENTIAL.
 FT CHAIN 40 2321 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 FT CHAIN 1629 2321 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1662 2321 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 40 1643 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1644 1664 POTENTIAL.
 FT DOMAIN 1665 2321 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 40 77 EGF-LIKE 1.
 FT DOMAIN 78 118 EGF-LIKE 2.
 FT DOMAIN 119 156 EGF-LIKE 3.
 FT DOMAIN 158 195 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 197 234 EGF-LIKE 5.
 FT DOMAIN 236 272 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 274 312 EGF-LIKE 7.
 FT DOMAIN 314 350 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 351 389 EGF-LIKE 9.
 FT DOMAIN 391 429 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 431 467 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 469 505 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 507 543 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 545 580 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 582 618 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 620 655 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 657 693 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 695 730 EGF-LIKE 18.
 FT DOMAIN 734 770 EGF-LIKE 19.
 FT DOMAIN 771 808 EGF-LIKE 20.
 FT DOMAIN 810 847 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 849 885 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 887 922 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).

Query Match 14.4%; Score 974; DB 1; Length 2321;

Best Local Similarity 25.0%; Pred. No. 1.2e-50;

Matches 304; Conservative 93; Mismatches 282; Indels 537; Gaps 70;

QY 59 SCTDILNWFKCTRHRVSYRTAYRHGEKTMYYRRKSQCCPGFYESGEMCVPHCADKC----- 113
 :| | :| : | || | : :| : | :|
 Db 250 TCVDGVNTYNC-----QCPPEW--TGQFCTED-VDECQLQPN 283

 QY 114 -VH--GRC---IAPNTCQCEPGWGGTNCSSACDG-----DHWGPHCTSR-----CQC--- 154
 | | | : :| | || | :|| | | | | | |
 Db 284 ACHNGGTCFNTLGGHSCVCVNGWTGESCSQNIDDCATAVCFHGATCHDRVASFYCACPMG 343

 QY 155 KNGALC-----NPITG--ACHCAAGFRGWRCE---DRCE----- 183
 | | || || : | | | | | : | |
 Db 344 KTGLLCHLDDACVSNPCHEDAICDTNPVNGRAICTCPPGFTGGACDQDVDECSIGANPCE 403

Qy 184 -----QGTYGNDCHQ-----RCQ-----CQNGATCDHVTGE--CRCPPGYTG 218
 ||:: | : ||: |:| ||| |: | | |:|
 Db 404 HLGRCVNTQGSFLCQCGRGYTGPRCETDVNECLSGPCRNQATCLDRIGQFTCICMAGFTG 463
 Qy 219 AFCE---DLCPPGKHGPPQCEQRCPCQNGGVC-HHVTG-ECSCPSGWMGTVC----- 264
 :|| | | | || ||||| | | |:||||: |: |
 Db 464 TYCEVDIDEC-----QSSPCVNGGVCKDRVNGFSCTCPSGFSGSTCQLDVDECAS 513
 Qy 265 -----GQP-----CPEGRFGKNCSQ---ECQ---CHNGGTCD-ATGQCHCSPG 301
 || | | | | : :| ||:| | | : | | :||
 Db 514 TPCRNGAKCVDQPDGYECRAEGFEGTLCDRNVDDCSPDPCHHGRCVDGFIASFSCACAPG 573
 Qy 302 YTGERCQDE-----CPVGTYGVLG-----AETCQ--- 325
 ||| ||: : || || || | : |
 Db 574 YTGTRCESQVDECERSQPCRHHGKCLDLVDKYLRCRPSGTTGVNCEVNIDDCASNPCFTGV 633
 Qy 326 CVNGGKCYHVSAGACLCEAGFAGERCEARL-----CPEGLYGIKCDKRCP--- 369
 | :| | | :||: || | | : | :| | :| ||
 Db 634 CRDGINRYD---CVCQPGFTGPLCNVEINECASSPCGEGGSCVDGENGFRC--LCPPGS 687
 Qy 370 ----CHLENTHSC-----HPMSGECACKPGWSGLYCNE----- 398
 | | :| | | | | ||:|||| |::
 Db 688 LPPLC-LPPSHPCAHEPCSHGICYDAPGGFRCVCEPGWSGPRCSQSLARDACESQPCRAG 746
 Qy 399 -----TCSPGFYGEACQOI--CS---CQNGADCDSVTGK---CTCAPGFKG-- 436
 || || | | : : | : | :| | :| | :| | :|
 Db 747 GTCSSDGMGFHCTCPPGVQGRQCELLSPCTPNPCEHGGRCESAPGQLPVCSCPQGWQGPR 806
 Qy 437 -----IDCSTPCPLGTYGINCSSRCCKNDVAVCSFVDGSCCTKAGWHGVDCSIRCPSGTW 491
 :|: | | | :|| |:: | |||| |: | |
 Db 807 CQQDVDEACAGPAPCGPHGI-CTNLAG-----SFSTCHGGYTGPSCDQDIND--- 852
 Qy 492 GFGCNLTQCQLNGGACNTLDG----TCTCAPGWRGEKC-----ELPCQDGTYGLNCA 539
 | : | ||||:| || :|| ||: | :| || || |
 Db 853 ---CDPN-PCLNGGSCQ--DGVSFSCSCLPGFAGPRCARDVDECLSNPCGPGT---CT 902
 Qy 540 ERCDCSHADGCHPTTGHCRCLPGWSGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGI 599
 : | : | | ||: | ||: | :|| | | || :| ||:
 Db 903 D-----HVASFTCTCPPGYGGFHCEQDL-----PDCS-PSSCFNGGTCV--DGV 943
 Qy 600 ----CECAPGFRGTTCQR-----ICSPGFYGHRC---SQTCPQC----- 631
 | | ||: | || :|| | || | | |||
 Db 944 NSFSCLCRPGYTGAHCQHEADPCLSRPCLHGGVCSAAHPGFRCTCLESFTGPQCQTLVDW 1003
 Qy 632 -----VHSSGPCHHITGLCDCLPGFTGALCNE----- 658
 : | | | | ||:| ||:
 Db 1004 CSRQPCQNGGRCVQTGAYCLCPPGWSGRLCDIRSLPCREAAAQIGVRLEQLCQAGGQCVD 1063
 Qy 659 -----VCPSGREGKNC---AGIC---TCTNNGTCNPI--DRSCQCYPGWIGSDC---- 699
 ||| || | :| | : ||| |:| ||: | :|
 Db 1064 EDSSHYCVCPEGRTGSHCEQEVDPCLAQPCQHGGFTRGYMGGYMCECLPGYNGDNCEDDV 1123
 Qy 700 ----SQPC-----PPAHWGPNCIHTCNCH 719
 |||| | | ||:| |
 Db 1124 DECASQPCQHGGSCIDLVARYLCSCPPGTLGVLCEINEDDCGPGPPLDSGPRCLHNGTCV 1183
 Qy 720 N--GAFCSAYDGECKCTPGWTGLYCTQ-----RCPLG 749

```

      : | |      | | ||: || |      | |
Db      1184 DLVGGF-----RCTCPPGYTGLRCEADINECRSGACHAAHTRDCLQDPGGGFRCLCHAG 1237
Qy      750 FYGKDCALI---CQ---CQNGADCDHISG-----QCTCRTGFMGRHCEQ----- 787
      | | | : | : ||: | | |      | | | | :
Db      1238 FSGPRCQTVLSPCESQPCQHGGQCRPSPGPGGGLTFTCHCAQPFWGPRCERVARSCRELQ 1297
Qy      788 -----KCPSTYGYGCRQI-----CDCLNNSTCDHIT--- 814
      | | | | |      | | : |
Db      1298 CPVGVPCQQTFRGPRCACPPGLSGPSCRSFPGSPPGASNASCAAAPCLHGGSCRPAPLAP 1357
Qy      815 -GTCYCSPGWKGARCD 829
      | | : || | ||:
Db      1358 FFRCAQGWGTGPRCE 1373

```

RESULT 12

NTC3_RAT

```

ID      NTC3_RAT          STANDARD;          PRT;  2319 AA.
AC      Q9R172;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN      NOTCH3.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Haritunians T., Boulter J., Weinmaster G., Schanen N.C.;
RT      "Rattus norvegicus mRNA for Notch 3.";
RL      Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN      [2]
RP      FUNCTION.
RX      MEDLINE=21094508; PubMed=11182080;
RA      Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA      Honjo T.;
RT      "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT      neural progenitor cells to an astroglial fate.";
RL      Neuron 29:45-55(2001).
RN      [3]
RP      TISSUE SPECIFICITY.
RX      MEDLINE=21331789; PubMed=11438922;
RA      Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT      "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT      functional roles for the Notch-DSL signaling system during brain
RT      development.";
RL      J. Comp. Neurol. 436:167-181(2001).
CC      -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC      Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC      Upon ligand activation through the released notch intracellular
CC      domain (NICD) it forms a transcriptional activator complex with
CC      RBP-J kappa and activates genes of the enhancer of split locus.
CC      Affects the implementation of differentiation, proliferation and
CC      apoptotic programs (By similarity). Acts instructively to control

```

CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within
 CC numerous cells throughout the CNS. It is more highly localized
 CC to ventricular germinal zones.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 34 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF164486; AAD46653.2; -.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 33.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 20.

DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_CA; 16.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 40 POTENTIAL.
 FT CHAIN 41 2319 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 FT CHAIN 1631 2319 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1664 2319 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 41 1645 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1646 1666 POTENTIAL.
 FT DOMAIN 1667 2319 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 41 79 EGF-LIKE 1.
 FT DOMAIN 80 120 EGF-LIKE 2.
 FT DOMAIN 121 158 EGF-LIKE 3.
 FT DOMAIN 160 197 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 199 236 EGF-LIKE 5.
 FT DOMAIN 238 274 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 276 314 EGF-LIKE 7.
 FT DOMAIN 316 352 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 353 391 EGF-LIKE 9.
 FT DOMAIN 393 431 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 433 469 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 471 507 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 509 545 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 547 582 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 584 620 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 622 657 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 659 695 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 697 732 EGF-LIKE 18.
 FT DOMAIN 736 772 EGF-LIKE 19.
 FT DOMAIN 773 810 EGF-LIKE 20.
 FT DOMAIN 812 849 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 851 887 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 889 924 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 926 962 EGF-LIKE 24.
 FT DOMAIN 964 1000 EGF-LIKE 25.
 FT DOMAIN 1002 1036 EGF-LIKE 26.
 FT DOMAIN 1038 1084 EGF-LIKE 27.
 FT DOMAIN 1086 1122 EGF-LIKE 28.
 FT DOMAIN 1124 1160 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1162 1205 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1207 1246 EGF-LIKE 31.
 FT DOMAIN 1248 1289 EGF-LIKE 32.
 FT DOMAIN 1291 1327 EGF-LIKE 33.
 FT DOMAIN 1337 1375 EGF-LIKE 34.
 FT REPEAT 1384 1420 LIN/NOTCH 1.
 FT REPEAT 1425 1461 LIN/NOTCH 2.
 FT REPEAT 1469 1503 LIN/NOTCH 3.

FT	REPEAT	1840	1869	ANK 1.
FT	REPEAT	1873	1903	ANK 2.
FT	REPEAT	1907	1936	ANK 3.
FT	REPEAT	1940	1969	ANK 4.
FT	REPEAT	1973	2002	ANK 5.
FT	SITE	1573	1574	CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY).
FT	DISULFID	44	56	BY SIMILARITY.
FT	DISULFID	50	67	BY SIMILARITY.
FT	DISULFID	69	78	BY SIMILARITY.
FT	DISULFID	84	95	BY SIMILARITY.
FT	DISULFID	89	108	BY SIMILARITY.
FT	DISULFID	110	119	BY SIMILARITY.
FT	DISULFID	125	136	BY SIMILARITY.
FT	DISULFID	130	146	BY SIMILARITY.
FT	DISULFID	148	157	BY SIMILARITY.
FT	DISULFID	164	176	BY SIMILARITY.
FT	DISULFID	170	185	BY SIMILARITY.
FT	DISULFID	187	196	BY SIMILARITY.
FT	DISULFID	203	214	BY SIMILARITY.
FT	DISULFID	208	224	BY SIMILARITY.
FT	DISULFID	226	235	BY SIMILARITY.
FT	DISULFID	242	253	BY SIMILARITY.
FT	DISULFID	247	262	BY SIMILARITY.
FT	DISULFID	264	273	BY SIMILARITY.
FT	DISULFID	280	293	BY SIMILARITY.
FT	DISULFID	287	302	BY SIMILARITY.
FT	DISULFID	304	313	BY SIMILARITY.
FT	DISULFID	320	331	BY SIMILARITY.
FT	DISULFID	325	340	BY SIMILARITY.
FT	DISULFID	342	351	BY SIMILARITY.
FT	DISULFID	357	368	BY SIMILARITY.
FT	DISULFID	362	379	BY SIMILARITY.
FT	DISULFID	381	390	BY SIMILARITY.
FT	DISULFID	397	410	BY SIMILARITY.
FT	DISULFID	404	419	BY SIMILARITY.
FT	DISULFID	421	430	BY SIMILARITY.
FT	DISULFID	437	448	BY SIMILARITY.
FT	DISULFID	442	457	BY SIMILARITY.
FT	DISULFID	459	468	BY SIMILARITY.
FT	DISULFID	475	486	BY SIMILARITY.
FT	DISULFID	480	495	BY SIMILARITY.
FT	DISULFID	497	506	BY SIMILARITY.
FT	DISULFID	513	524	BY SIMILARITY.
FT	DISULFID	518	533	BY SIMILARITY.
FT	DISULFID	535	544	BY SIMILARITY.
FT	DISULFID	551	561	BY SIMILARITY.

Query Match 14.4%; Score 969.5; DB 1; Length 2319;
Best Local Similarity 25.1%; Pred. No. 2.2e-50;
Matches 303; Conservative 71; Mismatches 304; Indels 531; Gaps 62;

```
QY      107 PHCAD-----KCVHGRCIAPNTCQCEPGWGGTNCSSACDGDHWGPHCTSRCQCKNGA 158
          |||           :||:         ||||| | |   |   ||| | |   |:
Db       42 PPCLDGSPCANGGRCTHQQPSREAAACLPLGWVGERCQLE-DPCHSGP-CAGRGVCQSSV 99

QY      159 LCNPITGACHCAAGFRGWRCE--DRC-----EQGTYG----- 188
```

```

      :      : | | | | | | | |
Db    100 VAGVARFSCRCLRGFRGPDCLPDCPFCSSPCAHPGAPCSVGS DGRYACACPPGYQGRNCRS 159

Qy    189 --NDCHQRCQCQNGATCDHVTG--ECRCPPGYTGAFCEd---LCPGKGHPQCEQRCPCQ 241
      : : | | : | | | | | : | | |
Db    160 DIDECRAGASCRHGGTCINTPGSFHCLCPLGYTGLLCENPIVPCAPS-----PCR 209

Qy    242 NGGVCHH---VTGECSCPSGWMGTVCg---QPCPEGRFGKNCSQECQCHNGGTC--DAAT 293
      | | | | | : | : | | | | | | |
Db    210 NGGTCRQSSDVTYDCACLPGFEGQNCENVNDDCPGHR-----CLNGGTCVDGVNT 259

Qy    294 GQCHCSPGYTGERCQ---DECPVGTYGVLCAETCQCVNNGGKCYHVSG--ACLCEAGFAGE 348
      | | | : | : | | : | | | : | : |
Db    260 YNCQCPPEWTGQFCTEDVDECQL-----QPNACHNGGTCFNLLGGHSCVCVNGWTGE 311

Qy    349 RCEARL-----CPEGLYGIKC--DKRC---PCHLENTHSCH 379
      | : | | | : | | | : |
Db    312 SCSQNIDDCATAVCFHGATCHDRVASFYCACPMGKTGLLCHLDDACVSNPCHEDAICDTN 371

Qy    380 PMSGE--CACKPGWSGLYCNE----- 398
      | : | | | : | : |
Db    372 PVSGRAICTCPPGFTGGACDQDVDECSIGANPCEHLGRCVNTQGSFLCQCGRGYTGPRCE 431

Qy    399 -----TCSPGFYGEACQ-----QICSCQNGADC-DSVT 425
      | | | | : | | | | |
Db    432 TDVNECLSGPCRNQATCLDRIGQFTCICMAGFTGTGFCEVDIDECQSSPCVNGGVCKDRVN 491

Qy    426 G-KCTCAPGFKGIDC-----STPCPLGTY-----GINCSSRCG-----C-KNDA 462
      | | | | | | | | | | | | | | : |
Db    492 GFSCCTCPSGFSGSTCQLDVDECASTPCRNGAKCVDQPDGYEC--RCAEGFEGTLCERNVD 549

Qy    463 VCSP-----VDG----SCTCKAGWHGVDCS-----IR 485
      | | | | | | | : | : |
Db    550 DCSPDPCHHGRCVDGIASFSCACAPGYTGIRCESQVDECRSQPCRYGGKCLDLVDKYLCR 609

Qy    486 CPSGTWGFGCNLTC-QCLNG----GACNTLDG----TCTCAPGWRGEKCEL----- 527
      | | | | | : | : | | | | | : | : |
Db    610 CPPGTTGVNCEVNIDDCASNPTFTGVCR--DGINRYDCVCQPGFTGPLCNVEINECASSP 667

Qy    528 -----PCQDGTYGLNC-----AERCDCSHADGCHPTTG--HCRCLPGW 563
      | | | | : | | | | | | | |
Db    668 CGEGGSCVDGENGFHCLCPPGSLPPLCLPANHPCAHKPCSHG-VCHDAPGGFQCVCDPGW 726

Qy    564 SGVHCDSVCAEGRWGPNCSLPCYCKNGASCSPDDGI---CECAPGFRGTTCQRI----- 614
      | | | | : | : | : | | | | : | : |
Db    727 SGPRCSQSLA-----PDACESQPCQAGGTCT-SDGIGFHCTCAPGFQGHQCEVLSPECTPS 780

Qy    615 -----CSPGFYGHRCSTCPQCVHSS-----GPCHHITG--LCDCL 648
      | | : | | | : | : | | : | |
Db    781 LCEHGGHCESDPDQLTVCSPPGWQGPCRQQDVDECA GASPCGPHGTCTNLPGSFRCICH 840

Qy    649 PGFTGALCNE-----VCPSGRFGKNCA----- 670
      | : | | : | | | | |
Db    841 GGYTGPFCDQDIDDCDPNPNCLNGGSCQDGVGSFSCSCLSGFAGPRCARDVDECLSSPCGP 900

Qy    671 GICT-----CTNNGTCNPIDR----SCQCYPGWIG 696
      | | | | | : | | | | : |

```

Db 901 GTCTDHVASFTCTCPPGYGGFHCETDLLDCSPSSCFNGGTC--VDGVNSFSCLCRPGYTG 958
 Qy 697 SDC-----SQP-----CPPAHWGPNCIHTCN-----CHNGAF 723
 : | : | | | | | | | | | |
 Db 959 THCQYKVDPCFSRPCLHGGICNPTHSGFEC--TCREGFTGNQCQNPVDWCSQAPCQNGGR 1016
 Qy 724 CSAYDGECKCTPGWTG-----LYCTQR-----CPL 748
 | | | | : | | | :
 Db 1017 CVQTGAYCICPPEWSGPLCDIPSLPCTEAAAHMGVRLEQLCQAGGQCIDKDHSHYCVCP 1076
 Qy 749 GFY GKDC-----ALICQCQNGADCDHISG--QCTCRTGFMGRHCEQ----- 787
 | | | | : | | | | | : | | |
 Db 1077 GRMGSHCEQEVDPCTAQPCQHGGTCRGYMGGYVCECPTGYSGDSCEDDVDECASQPCQNG 1136
 Qy 788 -----KCPSTYGYGCRQICD-----CLNNSTCDHITG--TCYCS 820
 | | | | | | | | : | | | | |
 Db 1137 GSCIDLVAHYLCSCPPGTLGVLCEINEDDCGPGPSLDGLRCLHNGTCVDLVGGFRCNCP 1196
 Qy 821 PGWKGARCD 829
 | : | :
 Db 1197 PGYTGLHCE 1205

RESULT 13

NTC4_HUMAN

ID NTC4_HUMAN STANDARD; PRT; 2003 AA.
 AC Q99466; O00306; Q99458; Q99940; Q9H3S8; Q9UII9; Q9UIJ0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 DE (hNotch4).
 GN NOTCH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
 RC TISSUE=Placenta;
 RX MEDLINE=97311416; PubMed=9168133;
 RA Sugaya K., Sasanuma S.-I., Nohata J., Kimura T., Fukagawa T.,
 RA Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
 RT "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
 RT human counterpart gene of mouse proto-oncogene Int3.";
 RL Gene 189:235-244(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Bone marrow, and Heart;
 RX MEDLINE=98360091; PubMed=9693032;
 RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
 RA Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 RT sequence of the human NOTCH4 gene.";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.

RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant."
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor."
 RL Am. J. Pathol. 154:785-794(1999).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=Q99466-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q99466-2; Sequence=VSP_001406;
 CC Name=3;
 CC IsoId=Q99466-3; Sequence=VSP_001407;
 CC -!- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 28 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.

CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in position 1438 to 1463.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D63395; BAA09708.1; ALT_FRAME.
DR EMBL; D86566; BAA13116.1; -.
DR EMBL; U95299; AAC32288.1; -.
DR EMBL; U89335; AAC63097.1; -.
DR EMBL; AB023961; BAB20317.1; -.
DR EMBL; AB024520; BAA88951.1; -.
DR EMBL; AB024578; BAA88952.1; -.
DR HSSP; P08709; 1BF9.
DR Genew; HGNC:7884; NOTCH4.
DR MIM; 164951; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 26.
DR Pfam; PF00066; notch; 2.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBLOOD.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; 11.
DR PROSITE; PS00022; EGF_1; 28.
DR PROSITE; PS01186; EGF_2; 21.
DR PROSITE; PS50026; EGF_3; 28.
DR PROSITE; PS01187; EGF_CA; 9.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW Triplet repeat expansion; Alternative splicing.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 2003 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
FT CHAIN 1432 2003 NOTCH EXTRACELLULAR TRUNCATION
FT (BY SIMILARITY).
FT CHAIN 1467 2003 NOTCH INTRACELLULAR DOMAIN
FT (BY SIMILARITY).

FT	DOMAIN	24	1447	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1448	1468	POTENTIAL.
FT	DOMAIN	1469	2003	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	24	63	EGF-LIKE 1.
FT	DOMAIN	64	115	EGF-LIKE 2.
FT	DOMAIN	118	155	EGF-LIKE 3.
FT	DOMAIN	156	192	EGF-LIKE 4.
FT	DOMAIN	194	232	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	234	274	EGF-LIKE 6.
FT	DOMAIN	276	312	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	314	353	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	355	391	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	392	430	EGF-LIKE 10.
FT	DOMAIN	432	473	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	475	511	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	513	549	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	551	587	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	589	625	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	626	659	EGF-LIKE 16.
FT	DOMAIN	661	689	EGF-LIKE 17.
FT	DOMAIN	691	727	EGF-LIKE 18.
FT	DOMAIN	729	765	EGF-LIKE 19.
FT	DOMAIN	767	803	EGF-LIKE 20.
FT	DOMAIN	806	842	EGF-LIKE 21.
FT	DOMAIN	844	880	EGF-LIKE 22.
FT	DOMAIN	882	928	EGF-LIKE 23.
FT	DOMAIN	930	966	EGF-LIKE 24.
FT	DOMAIN	968	1004	EGF-LIKE 25.
FT	DOMAIN	1006	1044	EGF-LIKE 26.
FT	DOMAIN	1046	1085	EGF-LIKE 27.
FT	DOMAIN	1087	1126	EGF-LIKE 28.
FT	DOMAIN	1130	1171	EGF-LIKE 29.
FT	DOMAIN	1472	1476	POLY-ARG.
FT	REPEAT	1165	1212	LIN/NOTCH 1.
FT	REPEAT	1213	1246	LIN/NOTCH 2.
FT	REPEAT	1247	1286	LIN/NOTCH 3.
FT	REPEAT	1633	1665	ANK 1.
FT	REPEAT	1666	1698	ANK 2.
FT	REPEAT	1700	1732	ANK 3.
FT	REPEAT	1733	1765	ANK 4.
FT	REPEAT	1766	1798	ANK 5.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	80	BY SIMILARITY.
FT	DISULFID	74	103	BY SIMILARITY.
FT	DISULFID	105	114	BY SIMILARITY.
FT	DISULFID	122	133	BY SIMILARITY.
FT	DISULFID	127	143	BY SIMILARITY.
FT	DISULFID	145	154	BY SIMILARITY.
FT	DISULFID	160	171	BY SIMILARITY.
FT	DISULFID	165	180	BY SIMILARITY.
FT	DISULFID	182	191	BY SIMILARITY.
FT	DISULFID	198	211	BY SIMILARITY.

Query Match 14.2%; Score 959.5; DB 1; Length 2003;
Best Local Similarity 26.5%; Pred. No. 7.4e-50;

Matches 298; Conservative 64; Mismatches 324; Indels 437; Gaps 64;

```
Qy      94  CCPGFYESGEMCVPHCADKC-----VHGRCIAPNT---CQCEPGWGGTNCSSACDGDH 143
      | ||| :|| | | | | | | | | | | | | | | | |
Db      105 CLPGF--TGERCQAKLEDPCPPSFCSKRGRCHI QASGRPQCSCMPGWTGEQCQLR----- 157

Qy      144 WGPHTSRCQCKNGALCNPITG--ACHCAAGFRGWRCE---DRCEQG---TYGNDCHQ- 193
      |:: | || :| | | | | | | | | | | | | | | |
Db      158 --DFCSAN-PCVNGGVCLATYPQIQCHCPPGFEGHACERDVNECFQDPGPCPKGTSCHNT 214

Qy      194 -----RCQ-----CQNGATC-----DHVTGECRCPPGYTGAFCE 222
      ||: | | | | | | | | | | | | | | | |
Db      215 LGSFQCLCPVGQEGPRCEL RAGPCPPRGCSNGGTCQLMPEKDSTFHLCLCPPGFIGPDCE 274

Qy      223 -----DLCPGKHG-----PQCEQRCP--CQNGGVCHH 248
      | | | | | | | | | | | | | | | | |
Db      275 VNPDCVSHQCQNGGTCQDGLDTYTCLCPETWTGWDCSEDVDECETQGPPHCRNGGTCQN 334

Qy      249 VTG--ECSCPSGWMGTVCQP-----CPEGRFGKNCSQE- 280
      | | | | | | | | | | | | | | | | |
Db      335 SAGSFHCVCVSGWGGTSCEENLDDCIAATCAPGSTCIDRVGSFSCLCPPGRTGLLCHLED 394

Qy      281 -C---QCHNGGTC--DAATGQ--CHCSPGYTGERCQ---DECPVGTYGVLC AETCQCVNG 329
      | | | | | | | | | | | | | | | | |
Db      395 MCLSQPCHGDAQCSTNPLTGSTLCLCQPGYSGPTCHQDLDECLMAQQG-----PSPCEHG 449

Qy      330 GKCYHVSGA--CLCEAGFAGERCEAR-----LCPEGLYGIK 363
      | | : | : | | | | : | | | | | | | | | |
Db      450 GSCLNTPGSFNCLCPPGYTGSRCEADHNECLSQPCHPGSTCLDLLATFHCLCPPGLEGQL 509

Qy      364 CD---KRC---PCHLENTHSCHPMSSG--ECACKPGWSGLYCNE----- 398
      | : | | | | | | | | | | | | | | | |
Db      510 CEVETNECASAPC--LNHADCHDLLNGFQCICLPGFSGTRCEEDIDECRSSPCANGGQCQ 567

Qy      399 -----TCSPGFYGEACQ---QIC---SCQNGADCDSVTGK--CTCAPGFKGIDCSTP 442
      | | | | | | | | | | | | | | | | |
Db      568 DQPGAFHCKCLPGFEGPRCQTEVDECLSDPCPVGASCLDLPGAFFCLCPSGFTGQLCEVP 627

Qy      443 -----CPLGTYGI-----NCSSRCG-CKNDAVCSPVDGSCTC 473
      | | | : | | | : | | : | | | | | |
Db      628 LCAPNLCQPKQICKDQKDKANCLCPDGSPGCAPPEDNCTCHHGHCQR-----SSCVC 679

Qy      474 KAGWHGVDC-----SIRCPSGTWGFGCN---LTCQ---CL 502
      | | | : | | | : | | | : | | | | | |
Db      680 DVGWTGPECEAELGGCISAPCAHGGTCYPQPSGYNCTCPTGYTGPTCSEEMTACHSGPCL 739

Qy      503 NGGACNTLDG--TCTCAPGWRGEKCE-----LPC-----QDGYTGLNCA----- 539
      | | | : | | | | | | | | | | | | | | | |
Db      740 NGGSCNPSPGGYYCTCPPSHTGPPCQTSTDYCVSAPCFNGGTCVNRPGTFSCLCAMGFQG 799

Qy      540 ERCD-----CSHADGCH--PTTGHCRCLPGWSGVHCHDS----- 570
      | | : | : | | | | | | : | | | | |
Db      800 PRCEGKLRPSCADSPCRNRATCQDSPQGPRCLCPTGYTGSGCQTLMDLCAQKPCPRNSHC 859

Qy      571 -----VCAEGRWGPNCSLP-----CYCKNGASCSPDDG---ICEC 602
      : | : | | | : | | | | | | | | | |
Db      860 LQTGPSFHCLCLQGWGTGPLCNLPLSSCQKAALSQGIDVSSLCHNGGLC-VDSGPSYFCHC 918
```

Qy 603 APGFRGTTTCQR-----ICSPGFYGHRCSTCPQCVHSSGP 637
 |||:|: || | :|: | | : | |
 Db 919 PPGFQGSILCQDHVNPCESRPCQNGATCMAQPSGYLCQCAPGYDQGNCSKELDAC--QSQP 976

Qy 638 CHHITGLCDCLPGFTGALCNEVCPSGRFGKNCAG-----ICTCTNNGTCNPIDRS-- 687
 ||: | | | | | | | | | | | | | : : :
 Db 977 CHN-HGTCTPKPG--GFHC--ACPPGFVGLRCEGDVDECLDQPCHTGTAAACHSLANAFY 1031

Qy 688 CQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDGE-----CKCTPGWTGLYC 742
 ||| || | | | | | | :| | | | | | | :| | |
 Db 1032 CQCLPGHTGQWCEVEIDPCHSQP-----CFHGGTCEATAGSPLGFICHCPKGFEGPTC 1084

Qy 743 TQRCPLGFYKDCALICQCQNGADC-----DHISGQCTCRTGFMGRHC-EQKCPSGTYG 795
 : | | ||: | :| | :| | :|: | | |
 Db 1085 SHRAPSCGFH-----HCHHGGLCLPSPKPGFPPRCACLSGYGGPDCLTPPAPK---- 1132

Qy 796 YGCRQICDCLNNSTCDHITG-----TCYCSPGWKGARCDQAG 832
 || | | | :| | | | | | :|
 Db 1133 -GCGPPSPCLYNGSCSETTGLGGPGFRCSCPHSSPGPRCQKPG 1174

RESULT 14

FBP1_STRPU

ID FBP1_STRPU STANDARD; PRT; 1064 AA.

AC P10079;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Fibropellin I precursor (Epidermal growth factor-related protein 1)
 DE (UEGF-1).

GN EGF1.

OS Strongylocentrotus purpuratus (Purple sea urchin).

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;

OC Strongylocentrotus.

OX NCBI_TaxID=7668;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90112459; PubMed=2514273;

RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;

RT "Structural analysis of the uEGF gene in the sea urchin

RT strongylocentrotus purpuratus reveals more similarity to vertebrate

RT than to invertebrate genes with EGF-like repeats.";

RL J. Mol. Evol. 29:314-327(1989).

RN [2]

RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.

RX MEDLINE=87319677; PubMed=3498216;

RA Hursh D.A., Andrews M.E., Raff R.A.;

RT "A sea urchin gene encodes a polypeptide homologous to epidermal

RT growth factor.";

RL Science 237:1487-1490(1987).

RN [3]

RP AVIDIN-LIKE DOMAIN.

RX MEDLINE=89196806; PubMed=2784773;

RA Hunt L.T., Barker W.C.;

RT "Avidin-like domain in an epidermal growth factor homolog from a sea

```

RT  urchin.";
RL  FASEB J. 3:1760-1764(1989).
RN  [4]
RP  CHARACTERIZATION.
RX  MEDLINE=91285254; PubMed=2060714;
RA  Bisgrove B.W., Andrews M.E., Raff R.A.;
RT  "Fibropellins, products of an EGF repeat-containing gene, form a
RT  unique extracellular matrix structure that surrounds the sea urchin
RT  embryo.";
RL  Dev. Biol. 146:89-99(1991).
CC  -!- FUNCTION: Form the apical lamina, a component of the extracellular
CC  matrix.
CC  -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC  OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC  THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC  EMBRYOS AND EARLY LARVAE.
CC  -!- ALTERNATIVE PRODUCTS:
CC  Event=Alternative splicing; Named isoforms=2;
CC  Name=IA;
CC  IsoId=P10079-1; Sequence=Displayed;
CC  Name=IB;
CC  IsoId=P10079-2; Sequence=VSP_000451;
CC  -!- DEVELOPMENTAL STAGE: Moderate levels in unfertilized eggs and
CC  during early cleavage, then rapidly increases in abundance between
CC  late morula and mesenchyme blastula stages to maximal levels
CC  maintained through subsequent stages. Expressed both maternally
CC  and zygotically.
CC  -!- SIMILARITY: Contains 21 EGF-like domains.
CC  -!- SIMILARITY: Contains 1 CUB domain.
CC  -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR TO
CC  AVIDIN/STREPTAVIDIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L08692; AAA62164.1; -.
DR  EMBL; L08692; AAA62163.1; -.
DR  EMBL; X17530; CAA35571.1; -.
DR  EMBL; M17421; AAA30050.1; -.
DR  EMBL; X17533; CAA35573.1; -.
DR  PIR; A40136; A40136.
DR  HSSP; P01132; 1EGF.
DR  InterPro; IPR000152; Asx_hydroxyl_S.
DR  InterPro; IPR005469; Avidin.
DR  InterPro; IPR005468; Avidin/str.
DR  InterPro; IPR000859; CUB.
DR  InterPro; IPR000742; EGF_2.
DR  InterPro; IPR001881; EGF_Ca.
DR  InterPro; IPR001438; EGF_II.
DR  InterPro; IPR006209; EGF_like.
DR  Pfam; PF01382; Avidin; 1.
DR  Pfam; PF00431; CUB; 1.

```

DR Pfam; PF00008; EGF; 21.
 DR PRINTS; PR00709; AVIDIN.
 DR PRINTS; PR00010; EGFBLOOD.
 DR SMART; SM00042; CUB; 1.
 DR SMART; SM00179; EGF_CA; 20.
 DR PROSITE; PS00010; ASX_HYDROXYL; 19.
 DR PROSITE; PS00577; AVIDIN; 1.
 DR PROSITE; PS01180; CUB; 1.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS01186; EGF_2; 19.
 DR PROSITE; PS50026; EGF_3; 21.
 DR PROSITE; PS01187; EGF_CA; 18.
 KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
 KW Glycoprotein; Calcium-binding.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1064 FIBROPELLIN I.
 FT DOMAIN 20 55 EGF-LIKE 1.
 FT DOMAIN 62 175 CUB.
 FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 860 896 EGF-LIKE 20.
 FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 936 1064 AVIDIN-LIKE.
 FT DISULFID 23 34 BY SIMILARITY.
 FT DISULFID 28 43 BY SIMILARITY.
 FT DISULFID 45 54 BY SIMILARITY.
 FT DISULFID 180 191 BY SIMILARITY.
 FT DISULFID 185 200 BY SIMILARITY.
 FT DISULFID 202 211 BY SIMILARITY.
 FT DISULFID 218 229 BY SIMILARITY.
 FT DISULFID 223 238 BY SIMILARITY.
 FT DISULFID 240 249 BY SIMILARITY.
 FT DISULFID 256 267 BY SIMILARITY.
 FT DISULFID 261 276 BY SIMILARITY.
 FT DISULFID 278 287 BY SIMILARITY.
 FT DISULFID 294 305 BY SIMILARITY.
 FT DISULFID 299 314 BY SIMILARITY.
 FT DISULFID 316 325 BY SIMILARITY.
 FT DISULFID 332 343 BY SIMILARITY.
 FT DISULFID 337 352 BY SIMILARITY.
 FT DISULFID 354 363 BY SIMILARITY.

FT	DISULFID	370	381	BY SIMILARITY.
FT	DISULFID	375	390	BY SIMILARITY.
FT	DISULFID	392	401	BY SIMILARITY.
FT	DISULFID	408	419	BY SIMILARITY.
FT	DISULFID	413	428	BY SIMILARITY.
FT	DISULFID	430	439	BY SIMILARITY.
FT	DISULFID	446	457	BY SIMILARITY.
FT	DISULFID	451	466	BY SIMILARITY.
FT	DISULFID	468	477	BY SIMILARITY.
FT	DISULFID	484	495	BY SIMILARITY.
FT	DISULFID	489	504	BY SIMILARITY.
FT	DISULFID	506	515	BY SIMILARITY.
FT	DISULFID	522	533	BY SIMILARITY.
FT	DISULFID	527	542	BY SIMILARITY.
FT	DISULFID	544	553	BY SIMILARITY.
FT	DISULFID	560	571	BY SIMILARITY.
FT	DISULFID	565	580	BY SIMILARITY.
FT	DISULFID	582	591	BY SIMILARITY.
FT	DISULFID	598	609	BY SIMILARITY.
FT	DISULFID	603	618	BY SIMILARITY.
FT	DISULFID	620	629	BY SIMILARITY.
FT	DISULFID	636	647	BY SIMILARITY.
FT	DISULFID	641	656	BY SIMILARITY.
FT	DISULFID	658	667	BY SIMILARITY.
FT	DISULFID	674	685	BY SIMILARITY.
FT	DISULFID	679	694	BY SIMILARITY.
FT	DISULFID	696	705	BY SIMILARITY.
FT	DISULFID	712	723	BY SIMILARITY.
FT	DISULFID	717	732	BY SIMILARITY.
FT	DISULFID	734	743	BY SIMILARITY.
FT	DISULFID	750	761	BY SIMILARITY.
FT	DISULFID	755	770	BY SIMILARITY.
FT	DISULFID	772	781	BY SIMILARITY.
FT	DISULFID	788	799	BY SIMILARITY.
FT	DISULFID	793	808	BY SIMILARITY.
FT	DISULFID	810	819	BY SIMILARITY.
FT	DISULFID	826	837	BY SIMILARITY.
FT	DISULFID	831	846	BY SIMILARITY.
FT	DISULFID	848	857	BY SIMILARITY.
FT	DISULFID	864	875	BY SIMILARITY.
FT	DISULFID	869	884	BY SIMILARITY.
FT	DISULFID	886	895	BY SIMILARITY.
FT	DISULFID	902	913	BY SIMILARITY.
FT	DISULFID	907	922	BY SIMILARITY.
FT	DISULFID	924	933	BY SIMILARITY.
FT	CARBOHYD	30	30	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	136	136	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSP LIC	477	780	Missing (in isoform IB).
FT				/FTId=VSP_000451.
FT	CONFLICT	279	279	L -> S (IN REF. 2).
SQ	SEQUENCE	1064 AA;	112072 MW;	2E569CA012ED6D09 CRC64;

Query Match 14.2%; Score 954.5; DB 1; Length 1064;
 Best Local Similarity 28.0%; Pred. No. 7.5e-50;
 Matches 290; Conservative 93; Mismatches 305; Indels 347; Gaps 69;

Qy 30 DPNVCSHWESYSVTVQESYPHPFDQIYYTSCDILNWFKCTRHRVSYRTAYRHGEKTMYR 89
 |||:| : :|||:| : ||
 Db 181 DPNLCQNG-----AACTDLVNDYACT----- 201
 Qy 90 RKSQCCPGFYESGEMC---VPHCA-DKCVHGRCIAPN---TCQCEPGWGGTNCSSACDG 141
 | ||| :| | : ||| | :| | ||:| | : :
 Db 202 ----CPPGF--TGRNCEIDIDECASDPCQNGGACVDGVNGYVCNCVPGFDGDECENNIN- 254
 Qy 142 DHWGPHTSRCQCKNGALCNPITGA----CHCAAGFRGWRCE---DRCEQGTYGNDCHQR 194
 | | | || :| : | | ||| | ||| | |
 Db 255 -----ECAS-SPCLNGGIC--VDGVNMFECTCLAGFTGVRCEVNIDECAS----- 296
 Qy 195 CQCQNGATC-DHVTG-ECRCPPGYTGAFCEDLCPGKGHPQCEQRCPQNGGVCHHVTGE 252
 |||| | | : | | || :|| ||: : :| | ||| | :
 Db 297 APCQNGGICIDGINGYTCSCLPGFSGDNCEN-----NDDECSS-IPCLNGGTCDVLVNA 349
 Qy 253 --CSCPSGWMGTVCQPCPEGRFGKNCSEQCQCHNGGTC-DAATG-QCHCSPGYTGERCQ 308
 | | || | | | : | ||| | | | | ||||| | :
 Db 350 YMCVCAPGWTGPTCADNIDE-----CA-SAPCQNGGVCIDGVNGYMCDCQPGYTGTHTCE 402
 Qy 309 ---DECPVGTYGVLCAETCQCVNNGGKCYH-VSG-ACLCEAGFAGERCEARLCPEGLYGIK 363
 ||| | ||| | :| :| || | ||
 Db 403 TDIDEC-----ARPPCQNGGDCVDGVNGYVCICAPGFDGLNCE----- 440
 Qy 364 CDKRCPCCHLENTHSC--HPMSGECACKPGWSGLYCNETCSPGFYGEACQ---QICS---C 415
 | | | | :| | ||| :| | : | :|
 Db 441 -----NNIDECASRPCQNGAVCVDGVNGFVC--TCSAGYTGVLCTDINECASMP 489
 Qy 416 QNGADC-DSVTGK-CTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDVACS-PVDG-SC 471
 || | | | | |||| ||:| :| | :| | :| | :| |
 Db 490 LNGGVCTDLVNGYICTCAAGFEGTNCETDTD-----ECAS-FPCQNGATCTDQVNGYVC 542
 Qy 472 TCKAGWHGVDC-----SIRCPSG-----TWGFGCNL-----TCQ----- 500
 || | : || | | :| | :| | ||:
 Db 543 TCVPGYTGVLCTDINECASFPCLNGGTCNDQVNGYVCVCAQDTSVSTCETDRDECASAP 602
 Qy 501 CLNGGAC-NTLDG-TCTCAPGWRGEKCEL-----PCQDGTYGLNCAERCDCSHADGC 550
 ||||| : :|| ||| ||| | ||: || :| || | : : | :
 Db 603 CLNGGACMDVNGFVCTCLPGWEGTNCIEINTDECASSPCMNG--GL-CVDQVN-SYV--- 655
 Qy 551 HPTTGHCRCCLPGWSGVHCDSDVCAEGRWGPNCSLPCYCKNGASC--SPDDGICECAPGFRG 608
 | ||||:|:| | : | | | | | | :
 Db 656 -----CFCLPGFTGIHCGTEIDECASSP-----CLNGGQCIDRVDSYECVCAAGYTA 702
 Qy 609 TTCQ-----RICSPGFYGHRCSTCPQCVHSSGPC----- 638
 || | :||:| | | :| :| ||
 Db 703 VRCQINIDECASAPCQNGGVCVDGVNGYVCNCAPGYTGDNCETEIDEC--ASMPCLNGGA 760
 Qy 639 --HHITG-LCDCLPGFTGALC-----NEVCPSGRFGKNCAGICTCTNNGTCNPIDRS 687
 : | | | :||| :| : | :| | ||
 Db 761 CIEMVNGYTCQCVAGYTGVICETDIDECASAPCQNG-----GVCTDTINGYI----- 807
 Qy 688 CQCYPGWIGSDCSQPCPPAHWGPNCIHTCNCHNGAFCSAYDG----ECKCTPGWTGLYCT 743
 | | ||: ||:| | | | | | | :| | :| ||
 Db 808 CACVPGFTGSNCETNIDECASDP-----CLNGGIC--VDGVNGFVCQCPPNYSPTYCE 858
 Qy 744 QRCPLGFYKDCALICQCQNGADCDHISGQ--CTCRTGFMGRHCE---QKCPSGTYGYGC 798


```

          |      | | | | | : :      | | : | : | : | : | : | : |
Db      859 -----ISLDACRSMPCQNGATCVNVGADYVCECVPGYAGQNCNEIDINECAS----- 904

Qy      799 RQICDCLNNSTC-DHITG-TCYCSPGWKGARCDQAGVIVGNL-----NSLSRTSTA 848
          | |      | | | | | | | : | : | : | : | : | : | : | : |
Db      905 ---LPCQNGGLCIDGIAGYTCQCRLGYIGVNCEEVGFCDLEGMWYNECNDQVTITKTSTG 961

Qy      849 LPADSYQIGAIAGIILVLVFLFLLALFIIYRHKQKGKESMPAVTYTPAMRVVNADYTI 908
          :      : | : | : | : | : | : | : | : | : | : | : |
Db      962 M-----MLGDYMTYNERALGYAAPTVVVVGYSN-----NYDFPS 995

Qy      909 SG-TLPHSNGGNANS 922
          | | :      | | : |
Db      996 FGFTVVRDNGQSTTS 1010

```

RESULT 15

NTC4_MOUSE

```

ID      NTC4_MOUSE      STANDARD;      PRT; 1964 AA.
AC      P31695; O35442; O88314; O88316; Q62389; Q62390; Q9R1W9; Q9R1X0;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Neurogenic locus notch homolog protein 4 precursor (Notch 4)
DE      [Contains: Transforming protein Int-3].
GN      NOTCH4 OR INT3 OR INT-3.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92194507; PubMed=1312643;
RA      Robbins J., Blondel B.J., Gallahan D., Callahan R.;
RT      "Mouse mammary tumor gene int-3: a member of the notch gene family
RT      transforms mammary epithelial cells.";
RL      J. Virol. 66:2594-2599(1992).
RN      [2]
RP      REVISIONS, SEQUENCE FROM N.A.
RX      MEDLINE=97294599; PubMed=9150355;
RA      Gallahan D., Callahan R.;
RT      "The mouse mammary tumor associated gene INT3 is a unique member of
RT      the NOTCH gene family (NOTCH4).";
RL      Oncogene 14:1883-1890(1997).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lung, and Testis;
RX      MEDLINE=96281668; PubMed=8681805;
RA      Uyttendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT      "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT      cell-specific mammalian Notch gene.";
RL      Development 122:2251-2259(1996).
RN      [4]
RP      SEQUENCE FROM N.A.
RA      Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA      Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RT      "Sequence of the mouse major histocompatibility locus class III

```

RT region.";
 RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracisternal type A particle-mediated activation of the Notch4/int3
 RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
 RT mRNAs by retroviral splicing events.";
 RL J. Virol. 73:5166-5171(1999).
 RN [6]
 RP FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uyttendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 RT Notch4 in embryonic endothelium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
 RN [7]
 RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 RP OF VAL-1463.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [8]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May regulate branching
 CC morphogenesis in the developing vascular system.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
 CC kidney, and at lower levels in the ovary and skeletal muscle. A
 CC very low expression is seen in the brain, intestine, liver and
 CC testis.
 CC -!- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
 CC embryonic development from 9.0 dpc.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme

CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: Loss of the extracellular domain causes constitutive
CC activation of the Notch protein, which leads to hyperproliferation
CC of glandular epithelial tissues and development of mammary
CC carcinomas.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 29 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M80456; AAB38377.1; -.
DR EMBL; U43691; AAC52630.1; -.
DR EMBL; U43691; AAC52631.1; -.
DR EMBL; AF030001; AAB82004.1; -.
DR EMBL; AB016771; BAA32281.1; ALT_SEQ.
DR EMBL; AB016772; BAA32283.1; ALT_INIT.
DR EMBL; AB016773; BAA32284.1; ALT_INIT.
DR EMBL; AB016774; BAA32285.1; -.
DR PIR; A38072; TVMVT3.
DR PIR; T09059; T09059.
DR HSSP; P08709; 1BF9.
DR MGD; MGI:107471; Notch4.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR000800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 27.
DR Pfam; PF00066; notch; 2.
DR PIRSF; PIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; 11.

DR PROSITE; PS00022; EGF_1; 28.
 DR PROSITE; PS01186; EGF_2; 21.
 DR PROSITE; PS50026; EGF_3; 27.
 DR PROSITE; PS01187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
 FT CHAIN 1411 1964 TRANSFORMING PROTEIN INT-3.
 FT CHAIN 1428 1964 NOTCH EXTRACELLULAR TRUNCATION.
 FT CHAIN 1463 1964 NOTCH INTRACELLULAR DOMAIN.
 FT DOMAIN 21 1443 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1444 1464 POTENTIAL.
 FT DOMAIN 1465 1964 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 21 60 EGF-LIKE 1.
 FT DOMAIN 61 112 EGF-LIKE 2.
 FT DOMAIN 115 152 EGF-LIKE 3.
 FT DOMAIN 153 189 EGF-LIKE 4.
 FT DOMAIN 191 229 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 231 271 EGF-LIKE 6.
 FT DOMAIN 273 309 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 311 350 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 352 388 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 389 427 EGF-LIKE 10.
 FT DOMAIN 429 470 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 472 508 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 510 546 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 548 584 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 586 622 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 623 656 EGF-LIKE 16.
 FT DOMAIN 658 686 EGF-LIKE 17.
 FT DOMAIN 688 724 EGF-LIKE 18.
 FT DOMAIN 726 762 EGF-LIKE 19.
 FT DOMAIN 764 800 EGF-LIKE 20.
 FT DOMAIN 803 839 EGF-LIKE 21.
 FT DOMAIN 841 877 EGF-LIKE 22.
 FT DOMAIN 878 924 EGF-LIKE 23.
 FT DOMAIN 926 962 EGF-LIKE 24.
 FT DOMAIN 964 1000 EGF-LIKE 25.
 FT DOMAIN 1002 1040 EGF-LIKE 26.
 FT DOMAIN 1042 1081 EGF-LIKE 27.
 FT DOMAIN 1083 1122 EGF-LIKE 28.
 FT DOMAIN 1126 1167 EGF-LIKE 29.
 FT REPEAT 1168 1208 LIN/NOTCH 1.
 FT REPEAT 1209 1242 LIN/NOTCH 2.
 FT REPEAT 1243 1282 LIN/NOTCH 3.

Query Match 14.1%; Score 951.5; DB 1; Length 1964;
 Best Local Similarity 24.9%; Pred. No. 2.2e-49;
 Matches 321; Conservative 75; Mismatches 348; Indels 545; Gaps 70;

Qy 95 CP-GFYESGEMCVPHCADKCV-----HGRCIAPNT----CQCEPGWGGTNCSSACDGDH 143
 || || :|: | | : | | | : | |||| | |
 Db 102 CPSGF--TGDRCQTHLEELCPPSFCSSNGGHCYVQASGRPQCSCEPGWTGEQCQLR----- 154
 Qy 144 WGPHTSRCQCKNGALCNPITG--ACHCAAGFRGWRCE-----DRCEQGTYGNDG 191

Db	155	--DFCSAN-PCANGGVCLATYPQIQRCPPGFEGHTCERDINECFLEPGPCPQGT---SC	208
QY	192	HQ-----RCQ-----CQNGATCD-----HVTGE-CRCPPGYTGA	219
Db	209	HNTLGSYQCLCPVQGEGPQCKLRKGACPPGSCLNNGGTCQLVPEGHSTFHLCLCPPGFTGL	268
QY	220	FCE-----DLCPPGKHG-----PQCEQRCP--CQNGGV	245
Db	269	DCMNPPDDCVRHQCCQNGATCLDGLDYYTCLCPKTKWGWDCSEDI DECEARGPPRCRNGGT	328
QY	246	CHHVTG--ECSCPSGWMGTVCQGP-----CPEGRFGKNCS	278
Db	329	CQNTAGSFHCVCVSGWGGAGCEENLDDCAAATCAPGSTCIDRVGSFSCCLCPPGRTGLLCH	388
QY	279	QE--C--QCHNGGTC--DAATGQ--CHCSPGYTGERCQ---DECPVGTGYVLCAETCQC	326
Db	389	LEDMLSQPCHVNAQCSTNPLTGSTLCICQPGYSGSTCHQDLDECQMAQQG-----PSPC	443
QY	327	VNGGKCYHVSGA--CLCEAGFAGERCEAR-----LCPEGLY	360
Db	444	EHGGSCINTPGSFNCLCLPGYTGSRCADHNECLSQPCHPGSTCLDLLATFHCLCPPGLE	503
QY	361	GIKCD---KRC---PCHLENTHSCHPMMSG--ECACKPGWSGLYCNE-----	398
Db	504	GRLCEVEVNECTSNPC--LNQAACHDLLNGFQCLCLPGFTGARCEKDMDECSSTPCANGG	561
QY	399	-----TCSPGFYGEACQQICS-----CQNGADCDSVTGK--CTCAPGFKGIDC	439
Db	562	RCRDQPGAFYCECLPGFEGPHCEKEVDECLSDPCPVGASCLDLPGAFFCLCRPGFTGQLC	621
QY	440	STP-----CPLGTYGI-----NCSSRCG-----	457
Db	622	EVPLCTPNMCQPGQCCQGEHRAPCLCPDGSPGCVPAEDNCPCHHGHQCQSLCVCDEGWT	681
QY	458	-----CKNDAVCSPVDG--SCTCKAGWHGVDCS---IRCPSGTW-----	491
Db	682	GPECETELGGCISTPCAHHGGTCHPQPSGYNCTCPAGYMGLTCSEEVTAACHSGPCLNNGGSC	741
QY	492	-----GFGCN-----LTCQCLNGGACNTLDGT--CTCAPGWRGEKCE-	526
Db	742	SIRPEGYSCTCLPSHTGRHCQTAVDHCVSASCLNNGGTCVNKPGTFFCLCATGFQGLHCEE	801
QY	527	-----LPCQDGTGYGLNC-----AERCDCSHADGCHPT--	553
Db	802	KTNPSCADSPCRNKATCQDTPRGARCLCSPGYTGSSCQTLIDLCAKPCPHTARCLQSGP	861
QY	554	TGHCRCPLPGWSGVHCD-----SVCAEGRWGP	579
Db	862	SFQCLCLQGWGTGALCDFPLSCQMAAMSQGIEISGLCQNGGLCIDTGSSYFCRCPPGFQ GK	921
QY	580	NCS---LPCY---CKNGASCSPDDG--ICECAPGFRGTTCQRI-----	614
Db	922	LCQDNMNPCEPNPCHHGSTCVPQPSGYVCQCAPGYEQNC SKVLEACQSQPCHNHGTCTS	981
QY	615	-----CSPGFYGHRCSTCPQCV----HSSG--PCHHITG--LCDCLPGFTGALCN-	657

Db 982 RPPGFHCACPPGFVGLRCEGDVDECLDRPCHPSGTAACHSLANAFYCQCLPGHTGQRC EV 1041
 Qy 658 -----EV-----CPSGRFGKNCA-----GICTCTNNGTCNPI 684
 | : | | | | : | | | | |
 Db 1042 EMDLCQSQPCSNNGSCEITTGPPPGFTCHCPKGFEGPTCSHKALSCGIHHCHNGGLCLPS 1101
 Qy 685 DRS-----CQCYPGWIGSDCSQPCPPAHWGP--NCIHTCNCHNGAFCSAYDGECKCTPGW 737
 : | | | : | | | | | : | | | |
 Db 1102 PKPGSPPLCACLSGFGGPDCLTPPAPPGCGPPSPCLH-----NGTCTETPGL 1148
 Qy 738 --TGLYCTQRCPLGFYKDCALICQCQNGADCDHISGQCTCRTG----- 779
 | | | | | : | : | | |
 Db 1149 GNPGFQCT--CPPDSPGPR----CQRPASGCEGRGGDTCDAGCSGPGGDWDGGDCSLG 1202
 Qy 780 -----FMGRHCEQKCPSGTY---GYGCR-----QIC-DCLNNS 808
 | | : | | | | | : |
 Db 1203 VPDPWKGCPPHSQCWLLFRDGRCHPQCDSEECLFDGYDCEIPLTCIPAYDQYCRDHFHNG 1262
 Qy 809 TCDHITGTCYCSPGWKGARCDQAGVIIVGNLNSLSRTSTAL-----PADSYQIGAIAGI 862
 | : | | | | : | | | | | : | : | :
 Db 1263 HCEKGCNNAEC--GWDGGDCRPEGE-----DSEGRPSLALLVVLRRPPALDQQLLALARV 1314
 Qy 863 IILVLVVLFLALFIIYRHKQKGKESSMP 891
 : | | | : : | : : |
 Db 1315 LSLTLRV---GLWV--RKDSEGRNMVFP 1337

Search completed: March 26, 2004, 16:09:48
 Job time : 39.1541 secs